Classification of complex networks in spatial, topological and information theoretic domains

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

d o c t o r r e r u m n a t u r a l i u m
(Dr. rer. nat.)
im Fach Physik
Spezialisierung: Theoretische Physik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
Humboldt-Universität zu Berlin

von
Marc Wiedermann, M. Sc.

Präsidentin der Humboldt-Universität zu Berlin:
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät:
Prof. Dr. Elmar Kulke

Gutachter:

1. Prof. Dr. Dr. h.c. mult. Jürgen Kurths
2. Prof. Dr. Holger Lange
3. Prof. Dr. Jörn Davidsen

Tag der mündlichen Prüfung: 1. Dezember 2017

Abstract

Complex network theory has in the past been proven as a powerful tool to quantify the structure of many real-world complex systems and its associated characteristics allow to discriminate such systems into different classes. In addition, so-called functional networks where links indicate functional or statistical interdependencies between dynamics at individual nodes have been utilized extensively to study, among other systems, the global climate system. In its *first* part, this work demonstrates the discriminative power of complex network theory to classify Eastern and Central Pacific phases of El Niño which were in the past not always consistently distinguished from each other using methods from classical statistical climatology. Therefore, an index based on evolving climate networks is proposed that objectively discriminates between both types. It confirms recent unanimously defined classifications of El Niño phases and also assigns types to the formerly ambiguous cases. In contrast to classical tools, the proposed index provides a meaningful discrimination into two types for La Niña episodes as well. After a thorough investigation of the climatic impacts of the thus discriminated flavors of El Niño and La Niña, this work moves from the classification of sets of single-layer networks to the more general study of interacting networks. Here, subnetworks represent oceanic and atmospheric variability and links between them indicate interactions between the two climatic subsystems. It is revealed that the ocean-to-atmosphere interaction in the Northern hemisphere follows a hierarchical structure and macroscopic network characteristics discriminate well different parts of the atmosphere with respect to their interaction with the ocean. Acknowledging that climate networks and a variety of real-world networks are in fact spatial networks, the *second* part of this work assesses the effect of the nodes' spatial embedding on the networks' topological characteristics, an effect that has so far been rarely taken into account. A hierarchy of null models is proposed which generate random surrogates from a given network such that global and local statistics associated with the spatial embedding are preserved. The proposed models capture macroscopic properties of the studied spatial networks much better than standard random network models and depending on the models' actual performance networks can ultimately be categorized into different classes. This thesis closes with extending the zoo of network classifiers by a two-fold metric to discriminate different classes of networks based on assessing their complexity. The basis of this metric is formed by a statistical complexity measure composed of a network's averaged per-node entropic measure and the associated Jenson-Shannon divergence. Within this framework networks of the same category tend to cluster in distinct areas of the resulting complexity-entropy plane. In particular, connectome networks exhibit among the highest complexity while, e.g., transportation and infrastructure networks display significantly lower values. The proposed framework further allows to objectively construct climate networks such that the statistical network complexity is maximized. In summary, the present work enhances the understanding of functional climate networks and provides valuable complements to existing methods from classical statistics. Secondly, the proposed classification schemes are generally applicable to study complex networks in many fields, ranging from social systems over infrastructures to neurophysiology.

Zusammenfassung

Die Theorie komplexer Netzwerke hat sich in der Vergangenheit als eine wirksame Methode bewiesen, um die Struktur vieler realer Systeme zu beschreiben. Die im Rahmen dieser Theorie entwickelten Methoden ermöglichen insbesondere eine Unterscheidung derartiger Systeme in verschiedene Klassen. Darüber hinaus wurden sogenannte funktionale Netzwerke intensiv genutzt, um unter anderem das globale Klimasystem zu untersuchen. In diesem Kontext bezeichnen Netzwerkkanten funktionale oder statistische Zusammenhänge zwischen Dynamiken an den individuellen Knoten des Netzwerks. Der *erste* Teil dieser Arbeit zeigt die Diskriminanzfähigkeit der Netzwerktheorie, um im Speziellen die Ost- und Zentralpazifischen Phasen des El Niño, welche in der Vergangenheit mitunter inkonsistent durch Methoden der klassischen statistischen Klimatologie unterschieden wurden, zu klassifizieren. In dieser Arbeit wird daher ein Index zur Klassifizierung beider Typen eingeführt, welcher auf der Evaluation sich zeitlich entwickelnder Klimanetzwerke basiert. Dieser Index bestätigt zum einen vergangene einstimmige Klassifikationen der zwei El Niño Phasen, zum anderen ordnet er die vormals mehrdeutigen Fällen einem der beiden Typen zu. Darüber hinaus und im Gegensatz zu klassischen Methoden bietet der vorgestellte Index die Möglichkeit einer Unterscheidung von La Niña in zwei Typen. Nach einer ausführlichen Studie der klimatischen Einflüsse der unterschiedenen Phasen von El Niño und La Niña verlegt die vorliegende Arbeit ihren Schwerpunkt vom Studium einer einzelnen klimatischen Schicht auf den generelleren Fall interagierender Klimanetzwerke. Im vorliegenden Fall repräsentieren einzelne Teilnetzwerke die jeweilige Variabilität in Ozean und Atmosphäre und die dazwischen liegenden Kanten bezeichnen Wechselwirkungen der beiden klimatischen Teilsysteme. Aus der Analyse dieser gekoppelten Klimanetzwerke resultiert die Erkenntnis, dass die Ozean-Atmosphären-Wechselwirkung einer hierarchischen Struktur folgt. Darüber hinaus erweisen sich makroskopische Netzwerkmaße als geeignet, um einzelne Atmosphärenschichten bezüglich ihrer Wechselwirkung mit dem Ozean zu unterscheiden. Da Klimanetzwerke sowie eine Vielzahl realer Netzwerke räumlich eingebettet sind, widmet sich der *zweite* Teil dieser Arbeit zunächst dem Einfluss dieses bisher nur selten berücksichtigten Effekts auf messbare topologische Netzwerkeigenschaften. In diesem Rahmen wird ein System von Nullmodellen eingeführt welche zufällige Surrogate eines gegebenen Netzwerks erzeugen, sodass bestimmte globale und lokale räumliche Eigenschaften des Systems erhalten bleiben. Die vorgeschlagenen Modelle erfassen die makroskopischen Eigenschaften der hier studierten Netzwerke deutlich besser als bisherige Standardmodelle zur Erzeugung von Zufallsnetzwerken. Abhängig von der Performanz der vorgeschlagenen Modelle können gegebene Netzwerke weiterhin in unterschiedliche Klassen eingeteilt werden. Diese Arbeit schließt mit einer Erweiterung der bisherigen Netzwerkklassifikatoren um eine zweidimensionale Metrik, welche verschiedene Klassen von Netzwerken auf Basis ihrer Komplexität unterscheidet. Diese Metrik basiert auf einem statistischen Komplexitätsmaß, welches aus einer gemittelten, knotenweise definierten Entropie sowie der zugehörigen Jenson-Shannon Divergenz besteht. Daraus ergibt sich, dass innerhalb dieser Methodik Netzwerke des gleichen Typs dazu neigen, in individuellen Bereichen der resultierenden Komplexitäts-Entropie-Ebene zu liegen. Insbesondere zeigen Konnektome mitunter

die höchste Komplexität, während beispielsweise Transport- und Infrastrukturnetzwerke deutlich geringere Werte aufweisen. Die vorgestellte Methode ermöglicht außerdem die objektive Konstruktion von Klimanetzwerken indem entsprechende Schwellwerte so gewählt werden, dass die statistische Komplexität maximiert wird. Die vorliegende Arbeit erweitert das Verständnis funktionaler Netzwerke und bietet Ergänzungen zu bestehenden Methoden der statistischen Klimatologie. Darüber hinaus sind die vorgestellten Klassifikationsschemen generell anwendbar um komplexe Netzwerke in sozialen Systemen, Infrastrukturen oder der Neurophysiologie zu studieren.

List of publications

This dissertation is partly based on the following publications. The identifiers, *e.g.*, P1, given below are cited in the text to highlight passages that are connected to one or more of these papers.

P13 (in review) M. Wiedermann, J. F. Donges, J. Kurths, and R. V. Donner. Mapping and discrimination of networks in the complexity-entropy plane. *preprint: arXiv:1704.07599*, 2017a

P12 (in review) M. Wiedermann, J. F. Siegmund, J. F. Donges, J. Kurths, and R. V. Donner. Differential imprints of distinct ENSO flavors in global extreme precipitation patterns. *preprint: arXiv:1702.00218*, 2017b

P11 (in review) P. P. Klamser, M. Wiedermann, J. F. Donges, and R. V. Donner. Zealotry Effects on Opinion Dynamics in the Adaptive Voter Model. *preprint: arXiv:1612.06644*, 2016

P10 W. Barfuss, J. F. Donges, M. Wiedermann, and W. Lucht. Sustainable use of renewable resources in a stylized social–ecological network model under heterogeneous resource distribution. *Earth System Dynamics*, 8(2):255–264, 2017. doi: 10.5194/esd-8-255-2017

P9 R. V. Donner, M. Wiedermann, and J. F. Donges. Complex network techniques for climatological data analysis. In C. L. Franzke and T. J. O'Kane, editors, *Nonlinear and Stochastic Climate Dynamics*. Cambridge University Press, 2017

P8 M. Wiedermann, J. F. Donges, D. Handorf, J. Kurths, and R. V. Donner. Hierarchical structures in Northern Hemispheric extratropical winter ocean-atmosphere interactions. *International Journal of Climatology*, 2016a. doi: 10.1002/joc.4956

P7 J. F. Siegmund, M. Wiedermann, J. F. Donges, and R. V. Donner. Impact of temperature and precipitation extremes on the flowering dates of four German wildlife shrub species. *Biogeosciences*, 13(19):5541–5555, 2016a. doi: 10.5194/bg-13-5541-2016

P6 M. Wiedermann, A. Radebach, J. F. Donges, J. Kurths, and R. V. Donner. A climate network-based index to discriminate different types of El Niño and

La Niña. *Geophysical Research Letters*, 43(13):2016GL069119, 2016c. doi: 10.1002/2016GL069119

P5 M. Wiedermann, J. F. Donges, J. Kurths, and R. V. Donner. Spatial network surrogates for disentangling complex system structure from spatial embedding of nodes. *Physical Review E*, 93(4):042308, 2016b. doi: 10.1103/PhysRevE.93.042308

P4 J. F. Donges, J. Heitzig, B. Beronov, M. Wiedermann, J. Runge, Q. Y. Feng, L. Tupikina, V. Stolbova, R. V. Donner, N. Marwan, H. A. Dijkstra, and J. Kurths. Unified functional network and nonlinear time series analysis for complex systems science: The pyunicorn package. *Chaos*, 25(11):113101, 2015a. doi: 10.1063/1.4934554

P3 M. Wiedermann, J. F. Donges, J. Heitzig, W. Lucht, and J. Kurths. Macroscopic description of complex adaptive networks coevolving with dynamic node states. *Physical Review E*, 91(5):052801, 2015. doi: 10.1103/PhysRevE.91.052801

P2 A. Rammig, M. Wiedermann, J. F. Donges, F. Babst, W. von Bloh, D. Frank, K. Thonicke, and M. D. Mahecha. Coincidences of climate extremes and anomalous vegetation responses: comparing tree ring patterns to simulated productivity. *Biogeosciences*, 12(2):373–385, 2015. doi: 10.5194/bg-12-373-2015

P1 D. C. Zemp, M. Wiedermann, J. Kurths, A. Rammig, and J. F. Donges. Node-weighted measures for complex networks with directed and weighted edges for studying continental moisture recycling. *Europhysics Letters*, 107(5):58005, 2014. doi: 10.1209/0295-5075/107/58005

Acknowledgements

I thank Prof. Jürgen Kurths for his continuous support, not only throughout the past three years of my PhD studies, but also during the time I spent at PIK to obtain my Bachelor's and Master's degree. I thank Reik Donner for his day-to-day supervision, the many ideas we developed, the studies we conducted together and for always allowing me to follow and pursue my own interests. I am also thankful for the financial support I received while being a part of his Young Investigators Group CoSy-CC^2 and I highly appreciate his efforts on funding my travels to as many destinations as possible. Specifically, I am grateful for being granted the opportunity to attend the 2014 Alpine Summer School on Dynamics, Stochastics and Predictability of the Climate System in Valsavaranche, the 2014 DAMES conference on Data analysis and modeling in Earth sciences in Milan, the 2015 and 2016 EGU General Assembly in Vienna, the 2015 NetSci International School and Conference on Network Science in Zaragoza and the 2016 Conference on Complex Systems in Amsterdam. Among many more, each of these schools and conferences broadened my horizon and allowed me to present and discuss my ongoing research with the scientific community. Beyond that, I am thankful for the opportunities to join multiple visits to the National Observatory of Athens and the Academy of Sciences of the Czech Republic in Prague. I also thank the German Academic Exchange Service (DAAD) for funding my trip to the 2016 AGU Fall Meeting in San Francisco where I was so fortunate as to present my work on the discrimination of different El Niño periods and its possibly related impacts on global extreme precipitation patterns.

I thank Jonathan Donges for his invaluable contributions to all joint papers that I was fortunate to seeing printed over the last three years. I always enjoyed the countless scientific discussions during our meetings over lunch and coffee as well as the occasional 'Feierabendbier'.

In addition, I am thankful to the entire CoSy team for making the past years such a memorable time, either by hanging around at the coffee machine or the scientific discussions during our (not always) weekly meetings. Specifically, I thank Jonatan Siegmund for being a great office mate, his efforts on making our workplace feel like a mixture of a living room and a botanical garden, and for the fun of working together on our joint papers.

Regarding my work on the assessment of spatial networks I thank Michael Gastner for providing his data on the airline, interstate, and Internet network. Further, Peter Menck thankfully provided his data on the Scandinavian power grid and I also thank

Sven Willner on behalf of the entire `zeean` team for providing the data on the world trade network. I further highly appreciate the chance to discuss my work on the statistical network complexity with Jobst Heitzig an Michael Small and I appreciate the great efforts by Prof. Aaron Clauset to compile the Colorado Index of Complex Networks.

I gratefully acknowledge the European Regional Development Fund (ERDF), the German Federal Ministry of Education and Research and the Land Brandenburg for supporting my work by providing resources on the high performance computer system at PIK. I thank the Cluster Support Team and especially Ciaron Linstead for fixing any kind of possibly emerging problem, sometimes even before I knew of its existence. Greatest thanks also goes to Norbert Marwan for setting up and maintaining PC74 which I used for many of the calculations presented in this work. Ultimately, I am very grateful for the time invested by all authors of various OpenSource Software packages that were used to compute and present the results that constitute the heart of this thesis.

I thank Benjamin Maier, Catrin Ciemer, Florian Klimm, Jakob Kolb, Jonathan Donges and Paul Schultz for iterating different parts of the manuscript at different stages of its development. Finally, greatest thanks goes to my parents, grandparents and family for their continuous moral support and for always encouraging me to follow my ideas.

Contents

List of Figures

List of Tables

List of frequently used mathematical symbols

$\mathcal{C}$	Global clustering coefficient
$\mathcal{H}$	Hamming distance
$\mathcal{L}$	Average path length
$\mathcal{T}$	Transitivity
$\mathcal{R}$	Rank of a matrix
$A_{\bullet\bullet}$	Entry of an adjacency matrix
C	Statistical network complexity
$D_{\bullet\bullet}$	Entry of a distance matrix
E	Set of links in a network
F	Cumulative distribution function
G	A network consisting of nodes and links
H	Entropy
K	Average degree of nodes in a network
M	Number of links in a network
N_x	Number of nodes in a network
N_t	Number of temporal sampling points in a time series
Q	Jenson-Shannon divergence
$R_{\bullet\bullet}$	Entry of a recurrence matrix
T	Threshold
V	Set of nodes in a network
$W_{\bullet\bullet}$	Entry of a weighted adjacency matrix
X	A time series
Y	An event series
d	Link density of a network
f	Probability distribution function
$k_\bullet$	Degree of a node in a network
l	Length of a link in a network
p	Probability
r	Radius
$s_\bullet$	Strength of a node in a network
t	Time
$w_\bullet$	Weight of a node in a network

List of frequently used mathematical symbols

w_{total}	Total weight of nodes in a network
λ	Latitudinal position
ν	Spatial dimension of a system
ϕ	Longitudinal position
π	Percentile
ρ	Pearson correlation coefficient

Chapter 1.

Introduction

Many real-world systems have been shown to be well described in terms of a complex network representation [Boccaletti et al., 2006, Newman, 2003]. Examples of these systems include social networks, such as herds of one or more species of animals [Croft et al., 2008, McGregor, 2005], transportation systems, such as road networks [Gastner and Newman, 2006, Kaiser and Hilgetag, 2004], or connectome networks, such as the human brain [Bullmore and Sporns, 2009, Sporns et al., 2005]. A complex network is usually comprised of two sets of fundamental entities, nodes and links. Nodes indicate a single entity of a complex system under study, such as individual animals or humans in social networks [Barnes and Burkett, 2010, Sundaresan et al., 2007], or intersections in road networks [Chan et al., 2011]. Links then indicate some interrelation between the nodes, such as friendship or proximity in the first and roads in the second of the above cases. It has been shown that such a representation of a given system in terms of nodes and links usually serves as a powerful tool to assess its complex structure [Albert and Barabási, 2002, Goh et al., 2002, Newman, 2010]. For the purpose of such a characterization of complex systems a variety of network characteristics has been defined which generally numerically quantify a complex network's topology in terms of the distribution of links between the corresponding nodes [Boccaletti et al., 2014, Costa et al., 2007, Newman, 2003]. Such metrics are usually employed to assess properties of the individual nodes or the entire network and, if desired, classify them into different categories [Albert and Barabási, 2002].

1.1. Classification of complex networks

As one prominent and intuitive example of a network classification problem, take an airline network, where nodes represent airports and links indicate scheduled flights between them [Bagler, 2008, Gastner and Newman, 2006, Guimerà et al., 2005]. Here, typically two types of nodes emerge. The first type, usually denoted as *hubs*, are nodes with a comparatively large number of links attached to them, i.e., many in- and outbound flights are scheduled at the corresponding airport. A second type of nodes, usually denoted as *peripherals*, are those with only few links attached

and, thus, indicating smaller remote airports with only few scheduled flights. The specific network characteristic that quantifies such a number of attached links to an individual node is denoted the node's *degree*. The distribution of the degree along all nodes in the network often follows a power-law [Barabasi, 2009, Barabási and Albert, 1999, Goh et al., 2002] in which case the network is called a *scale-free* network.

In addition to the above exemplary discrimination of nodes into distinct types, complex network theory also provides valuable tools to intercompare the macroscopic structure of different networks [Amaral et al., 2000, Latora and Marchiori, 2003]. In this case, it is often advisable to utilize so-called *global* network characteristics that reduce a networks' structure to a single number [Albert and Barabási, 2002]. The two most prominent examples of such global characteristics are probably the *clustering coefficient* that measures the fraction of closely connected triples of links or the *average path length* which gives the average number of steps one takes when traveling between nodes in the network [Newman, 2003]. These two metrics are then often utilized to estimate whether a given network obeys the *small-world* property [Amaral et al., 2000, Watts and Strogatz, 1998]. This particular property is on the one hand characterized by a strong tendency of the system to divide into distinct clusters of nodes, which correspond for example to different families or groups in a social network [Ebel et al., 2002]. On the other hand these groups must be sufficiently strongly connected such that pairs of nodes in the network are only separated from each other by a low number of steps. Specifically, the latter property has been observed in the early second half of the 20th century in a famous experiment conducted by Milgram [1967] which revealed that specifically humans are connected with only a few degrees of separation between them. Based on the assessment of the above mentioned global metrics, networks are then classified as small-world networks if their clustering coefficient is large while their average path length is comparatively low [Watts and Strogatz, 1998].

Both of the two stereotypical examples given above are meant to indicate that complex network theory is particularly useful for characterizing and discriminating certain roles of either distinct nodes within a network (as in the example of the airline network) or to intercompare the structure of different networks and assign unique types to each of them. In addition to the examples introduced above, other approaches have for example successfully distinguished between different classes of scale-free networks by means of characteristics associated with their degree distribution [Goh et al., 2002]. Furthermore, networks have been assigned to so-called superfamilies based on the distribution of certain motifs that form their substructure [Milo et al., 2004]. Another successful approach to quantify topological differences in networks of different types is based on examining their community structure [Fortunato, 2010], that is a grouping of the network into smaller closely connected sets of nodes. Depending on the specific application, such an approach has been shown to yield

statistical properties within the communities that are unique to different types of networks under study [Lancichinetti et al., 2010].

Building on these observations, the present work presents four consecutive applications that utilize and further develop methods from complex network theory to discriminate different stages of one system or classes of different systems under study. In particular, two distinct applications utilize functional climate network analysis [Donges et al., 2009b, Tsonis et al., 2006] to discriminate different temporal or spatial characteristics in the evolution of the climate system. The specifics of these analyses are outlined in Sec. 1.2 below and discussed in detail in Pt. I of this work.

From there, Part II of this work addresses in detail two major conceptual issues with regard to the classification of complex networks. First, the effect of a given network's spatial embedding on the assessment of metrics such as clustering coefficient and average path length is discussed in Chap. 7. Specifically, this work proposes a hierarchy of null models to correct for biases that might be induced into the assessment of the aforementioned metrics by the spatial embedding of the system. Ultimately, this work concludes with extending the zoo of previously defined network classifiers by means of a novel characteristic which allows to quantify the actual complexity of a given network and compare its values across different types of networks (Chap. 8). Both of these latter two applications are discussed with respect to a variety of real-world networks. Again, the specifics of these two approaches are outlined in Sec. 1.3 below and discussed in detail in Pt. II of this work.

1.2. Functional climate networks

In addition to the study of real-world systems, such as infrastructure or communication networks, complex network theory has in the past also been established as a powerful tool to study statistical interdependencies in the climate system [Donges et al., 2009b, 2015b, Tsonis and Roebber, 2004, Tsonis et al., 2006, 2008]. In these so-called *functional climate networks* nodes represent different grid points or measurement sites on or parallel to the Earth's surface at which some climatic data set (mostly in terms of individual time series) is available. Links then represent *functional* interdependencies usually indicated by, e.g., significant correlations [Donges et al., 2009a,b, Paluš et al., 2011, Radebach et al., 2013] or the synchronous occurrence of extreme events [Boers et al., 2013, 2015, Malik et al., 2010, 2011, Stolbova et al., 2014] between the time series that are provided at each node. As a complement to the assessment of such pairwise statistical interrelations, climate networks have also been successfully constructed from investigating causal interrelationships between different processes observed at the nodes [Runge et al., 2012a, 2013]. Such a network representation of the underlying system, also referred to as *causal effect*

networks [Kretschmer et al., 2016], specifically allows to filter out potentially misleading effects of auto-dependencies, common drivers, or indirect couplings [Ebert-Uphoff and Deng, 2012, Runge et al., 2012b, 2015]. In addition to studies on observational data of climate dynamics, climate networks have also been applied to, e.g., successfully hindcast extreme events, such as extreme precipitation in South America [Boers et al., 2014a], or to predict the occurrence of El Niño episodes [Ludescher et al., 2013, 2014].

The majority of studies conducted within the climate network framework so far focused solely on assessing the statistics of a single network composed of one climatic field or layer, such as surface air temperature [Donges et al., 2009a], sea level pressure [Tsonis and Roebber, 2004], or ocean temperature variability at the surface [Feng and Dijkstra, 2014, Tantet and Dijkstra, 2014] and different depths [van der Mheen et al., 2013]. In all these cases the focus was put on the detection of certain nodes or larger spatially connected sets thereof that obey some outstanding properties such as an increased betweenness centrality indicating for the presence of major oceanic currents [Donges et al., 2009a], or an increased node degree corresponding to key areas in forecasting Indian monsoon dynamics [Stolbova et al., 2014, 2016].

In addition to the extensively performed quantification of single node characteristics in climate networks, the present work explores the possibility to discriminate larger sets of various climate networks into different types and interpret these classifications as well as the involved metrics with respect to the specific application.

Different phases in a climate network's temporal evolution

The present work opens with an assessment of the El Niño Southern Oscillation (ENSO) which alternates between positive (El Niño) and negative (La Niña) phases [Trenberth, 1997]. ENSO has already been studied intensively using climate network analysis in order to quantify corresponding teleconnections [Gozolchiani et al., 2011, Tsonis and Swanson, 2008, Tsonis et al., 2008], its effect on other climatic subsystems [Gozolchiani et al., 2008] and the dynamics of its related oceanic wave dynamics [Wang et al., 2016]. Additionally, climate network approaches, allowed for successfully forecasting El Niño by assessing the strength of linkages in the equatorial Pacific [Ludescher et al., 2013, 2014].

Notably, especially the El Niño phase further exhibits two distinct types which are usually and in many cases not consistently discriminated by different spatial patterns of sea surface temperature (SST) anomalies [e.g., Ashok et al., 2007, Kao and Yu, 2009, Kug et al., 2009, Yeh et al., 2009]. The first type (the classic or Eastern Pacific (EP) El Niño [Harrison and Larkin, 1998, Rasmusson and Carpenter, 1982]) is characterized by strong positive SST anomalies close to the western coast of South America, while the second type (referred to as El Niño Modoki or Central Pacific (CP) El Niño) exhibits the strongest SST anomalies close to the dateline.

These different types have distinct impacts on the global climate system, such as an increase in rainfall over northern and eastern Australia during CP El Niños [Ashok et al., 2007, Taschetto and England, 2009] which is contrasted by a significant rainfall reduction over eastern Australia during EP El Niños [Chiew et al., 1998]. Because of these differential imprints and characteristics, a proper discrimination of the two El Niño (and possibly La Niña) types is key to assess its possible impacts on other climate subsystems.

Acknowledging the discrimination power of complex network theory and its associated metrics, this work introduces a method to distinguish the two different El Niño types based on the assessment and classification of sets of climate networks constructed for different time intervals [Radebach et al., 2013] representing the temporal evolution of the climate system under study. In particular, the proposed methods rely on the assessment of the climate network *transitivity*, a measure that is closely entangled with the clustering coefficient as introduced in Sec. 1.1. This metric provides a scalar-valued measure that quantifies the (disperse vs. strongly localized) spatial distribution of teleconnections [Tsonis et al., 2006] along the globe and is compared to a baseline value defined by the transitivity of networks computed from 30-year windows that are centered around the period of interest. The present work shows that the surpassing of that threshold defines an EP El Niño, while the opposite case indicates a CP El Niño. In comparison with recent studies, the proposed method confirms all EP and CP El Niños between 1951 and 2014 that were commonly defined by Hendon et al. [2009], Hu et al. [2011], Kim et al. [2011], Kug et al. [2009], Larkin and Harrison [2005], Yeh et al. [2009] and Graf and Zanchettin [2012] and provides a consistent assignment for those periods that were ambiguously classified so far.

The details of this study are discussed in Chap. 4 of this work. Possible differential impacts of the thus obtained discrimination of the different ENSO phases are then analyzed with a special focus on extreme precipitation responses in Chap. 5.

Classification of interactions between different climatic fields

Together with the proposed assessment and classification of different ENSO periods it is observed that most previous studies conducted within the framework of climate networks so far focused solely on the dynamics within a single climatic field or layer such as the ocean [Feng and Dijkstra, 2014, Tantet and Dijkstra, 2014, van der Mheen et al., 2013] or one atmospheric layer at or parallel to the Earth's surface [Boers et al., 2013, Stolbova et al., 2014, Tsonis et al., 2010]. It is however well established that the dynamics within, e.g., the two major subcomponents of the Earth's climate system, ocean and atmosphere, are closely entangled [Frankignoul et al., 2001, Trenberth and Hurrell, 1994]. Examples for these interactions include the North Atlantic eddy-driven jet stream [Woollings et al., 2010] or the Pacific ocean forcing to the atmosphere [Wyrtki, 1975].

Inspired by approaches to investigate the interaction structure between different mutually coupled subsystems such as infrastructure networks [Boccaletti et al., 2014, Buldyrev et al., 2010, Vespignani, 2010] a novel set of coupled network measures has been proposed by Donges et al. [2011b] which provides a general tool to quantify interdependencies between such subcomponents in complex coupled climate networks. The corresponding framework has already been successfully applied to investigate interactions between different layers of geopotential height fields and to study ocean-atmosphere interactions in the tropical Pacific [Feng et al., 2012] or over the South Atlantic Convergence Zone [Tirabassi et al., 2015].

Following the assessment of different ENSO periods by means of single-layer network statistics, this work follows the approach by Donges et al. [2011b] and presents an exploratory study to quantify and classify different spatial characteristics of ocean-atmosphere interactions by means of coupled climate network analysis. In particular, an investigation of associated per-node characteristics reveals that the statistical interrelation between ocean and atmosphere exhibits a hierarchical structure, in which individual parts or areas of the ocean surface connect with multiple statistically dissimilar parts of the atmosphere. In that sense, methods from coupled climate network distinguish well between spatial signatures in the two different subnetworks that are related with their mutual interaction. Building upon previous studies by, e.g., Czaja and Frankignoul [1999], Frankignoul and Sennéchael [2007] and Gastineau and Frankignoul [2015] the observed hierarchy may be related to dominant atmospheric patterns forcing the ocean on the time scales investigated in this study.

From there, a further assessment of global metrics that are specifically tailored to study the interaction between two individual climate networks proof to discriminate well between parts of the atmosphere, such as troposphere and stratosphere, in terms of topological characteristics related to their interaction with the ocean.

The specific details related to this study of coupled climate networks are presented in Chap. 6.

1.3. Structural (real-world) networks

In contrast to the application of complex network theory to climatological problems which commonly gives rise to the study of functional networks, the second part of this work is devoted to the extension of general theoretical approaches related to the discrimination of complex networks. At the same time, and to make the results comprehensive, the primary focus is shifted from the study of functional networks to that of real-world networks, such as social [Ebel et al., 2002], transportation [Chan et al., 2011], or communication networks [Gastner and Newman, 2006]. As for those cases links usually, and in contrast to the case of functional climate networks, directly correspond to physical or (at least) measurable connections between the entities

represented by the networks' nodes, these types of networks are often referred to as *structural* networks [Bullmore and Sporns, 2009, Newman, 2003].

Spatially embedded networks

To start, it is acknowledged that the climate networks that were discussed in Sec. 1.2 are in fact spatially embedded in a metric space formed by (a part of) the Earth's surface or parallel layers above [Tsonis et al., 2006]. It has recently been discovered that such spatially embedded networks must be treated with special care when assessing their corresponding topological characteristics [Heitzig et al., 2012]. In particular, one should account for the generally heterogeneous distribution of nodes along the domain of interest which in the case of climate networks is usually achieved by weighing nodes according to the cosine of their latitudinal position [Feng et al., 2012, Tsonis and Roebber, 2004, Wiedermann et al., 2013]. Such a weighing is sufficient as for most tasks of climate network analysis, and specifically for the ones discussed in this work, nodes are usually located on a regular grid of equal spacing in latitudinal and longitudinal direction. This yields the same number of nodes per latitude and latitude and thus a monotonous increase of the nodes' density from Equator to the Pole [Tsonis et al., 2006].

Generally, most real-world complex systems that exhibit a network structure are also spatially embedded in some metric space [Barthélemy, 2011]. Examples of such networks include again social [Capocci et al., 2006] as well as transportation and distribution [Banavar et al., 1999, Chan et al., 2011, Jarvis et al., 2015], communication [Buldyrev et al., 2010], or electricity networks [Buzna et al., 2009]. In real-world networks nodes are (if spatially embedded) generally distributed rather heterogeneously along the domain of interest making a comparative simple weighting scheme as employed for climate networks unfeasible. Think for example again of the airline network with airports scattered unevenly across the globe or road networks where roads follow some geographical and logistical constraints rather than a perfect grid [Bagler, 2008, Chan et al., 2011, Gastner and Newman, 2006]. It has been shown by Bialonski et al. [2010] that in many of such cases the assessment of global characteristics that are utilized to assign, e.g., the small-world property to a given network are largely effected by the corresponding spatial embedding. In other words, it often remains unquantified whether a certain categorization of a network is to some extent already explicable as emerging from the network's spatial embedding alone.

Addressing this issue, Chap. 7 of this work provides a framework to systematically and explicitly study the general influence of a network's spatial embedding on its resulting macroscopic characteristics. For this purpose a set of random network models to create surrogates that preserve certain geographical and topological features of a given network is proposed. The network surrogates are constructed by iteratively

rewiring the original network while preserving a set of its low-order geographical features. In particular, one model preserves, in addition to the degree sequence, the global link-length distribution while a second model additionally preserves for each node the length distribution of the links connected to it. The resulting surrogate networks allow for evaluating to what extent observed global properties of a given network are explicable by geometric constraints inflicted on the network's structure while no assumptions on the specific construction principles are necessary. Applying this framework to a variety of real-world networks, i.e., the US airline network, the US interstate network, the Internet [Gastner and Newman, 2006], the Scandinavian power grid [Menck et al., 2014], a world trade network [Lenzen et al., 2012], and the road network of a German city, reveals that the macroscopic properties of a certain set of networks are only reproduced by applying either of the two geometrically constrained models proposed in this work, while the consideration of topological features alone is not sufficient. This in turn implies, that the topological structure of the affected networks is to a large extent already properly explained by the spatial embedding of the corresponding systems alone.

Discrimination based on complexity

Up to this point characterizations and classifications of complex networks have been discussed solely with respect to their associated topological or spatial characteristics. However, in the past a third class of characteristics emerged that may be utilized to describe a complex network's topology by means of information theoretic approaches, e.g., in terms of an appropriately defined entropy measure [Anand and Bianconi, 2009, Dehmer and Mowshowitz, 2011, Konstantinova et al., 2003]. Interestingly, these metrics have often been promoted as quantifiers for the actual *complexity* [Bonchev and Buck, 2005, Dehmer et al., 2009] of a network, thus answering the often posed question: *How complex is a complex network?* [Anand and Bianconi, 2009]. However, in most cases these measures are tailored to specific applications and have so far not been successfully applied to intercompare different types or classes of networks as in this respect they often lack a meaningful interpretation [Dehmer et al., 2009].

Contributing to the above issues, this work ultimately proposes a two-dimensional metric based on an entropic and an adjoint statistical complexity measure to distinguish different types of complex networks. This approach was originally introduced to distinguish chaotic from stochastic systems in time series analysis without making any assumptions on the underlying process that generated the respective series [Martin et al., 2006, Rosso et al., 2007] and, thus, making it generally applicable to any kind of system. It assigns each system under study a position in a two dimensional space spanned by an entropy and a statistical complexity measure, where the latter is given as a product of entropy and Jenson-Shannon divergence with respect to a

uniform distribution.

Chap. 8 transfers this concept from time series to the case of complex networks and redefines the above entropy and statistical complexity accordingly. Specifically, the framework proposed here relies on one definition of entropy that is defined in a node-wise fashion and is based on the probability to jump from a specific node to its neighbors in the network [Small, 2013]. This notion of entropy is closely related to random walks which themselves are in their application and interpretation closely related with the assessment of a networks' navigability and thus, complexity [Amaral and Ottino, 2004, Barthélemy, 2011, White and Houseman, 2002].

An application of this formalism to 29 real-world networks that are discriminated by context into the four types of social animal, social affiliation, transportation and connectome networks reveals that in most cases the different types occupy distinct areas in the complexity entropy-plane. This implies that the proposed framework serves to naturally distinguish different types of systems under study by means of assessing their individual statistical complexity. Ultimately, linking back to Pt. I of this work, the proposed network complexity measure also proofs to be useful to objectively construct threshold-based networks, such as the functional climate networks discussed above [Donges et al., 2009b, Tsonis et al., 2006] or recurrence networks [Donner et al., 2010, 2011, Marwan et al., 2009] by choosing a discrete network representation that maximizes statistical complexity.

1.4. Structure of this thesis

The remainder of the present work is now organized as follows. Chap. 2 presents the necessary theoretical background of complex network theory that forms the basis to all following studies. From there, the works splits in two parts. Part I is devoted to the study and discrimination of functional climate networks. It opens in Chap. 3 with a general introduction to climate network analysis and associated methods thereof. Chap. 4 presents the objective classification of different El Niño and La Niña phases as outlined in Sec. 1.2 and the closely related Chap. 5 investigates and discusses possible impacts of the thus discriminated phases on global extreme precipitation patterns. Chap. 6 then moves from the case of single-layer to coupled climate networks and specifically includes, in addition to the atmospheric variability studied in Chap. 4, also ocean dynamics as well as its associated interactions with the atmosphere.

Part II then moves to the study of real-world network as well as the further development of general network theoretic frameworks. Chap. 7 discusses the effects of spatial embedding on the assessment of global network characteristics as one possible choice of network classifiers and introduces appropriate random network models to investigate these effects. Ultimately, Chap. 8 proposes a novel network classification

scheme based on assessing a network's actual *complexity* in terms of a two-fold information theoretic metric. Finally, the work is summarized and conclusions are drawn in Chap. 9.

Chapter 2.

Theoretical background

This chapter introduces the theoretical foundation that all results and derivations presented in this work are based on. Specifically, it opens with a general introduction to the common terminology that is utilized when discussing network theory. The notions of an adjacency matrix $\mathbf{A}$, weight matrix $\mathbf{W}$ and node weights $w_\bullet$ are introduced as a basis for the derivation of the complex network characteristics that are utilized for all analyses presented in this work. Particularly, the study of the structural real-world networks relies on classical unweighted network measures, such as the global clustering coefficient $\mathcal{C}$ and average path length $\mathcal{L}$. In turn, the functional climate networks presented in this work are studied and quantified by means of of node- and (in parts) link-weighted network characteristics. For the kind of functional climate networks considered here, the networks' nodes should generally be weighted according to the share of the area on or parallel to the Earth's Surface that they represent. Links should (whenever necessary and applicable) be weighted with respect to the strength of interdependency between two nodes.

All further methodological details with respect to the studies presented in this work that go beyond the notions of standard formalisms in complex network theory are discussed in the specific chapters that introduce the results of each study.

This chapter is partly based on the methodological sections of the works P5 [Wiedermann et al., 2016b], P6 [Wiedermann et al., 2016c]and P8 [Wiedermann et al., 2016a].

2.1. Network definition and nomenclature

A network $G = G(V, E)$ usually consists of a set V of N_x nodes that represent certain discrete entities of a system under study, such as single neurons in neural networks [Bullmore and Sporns, 2009], intersections in road networks [Chan et al., 2011], or geographic locations in climate networks [Donges et al., 2009b, Donner et al., 2017]. Nodes are usually uniquely labeled with an integer value $i = 1, \ldots, N_x$ and are connected via M links of set E that represent a physical or functional link

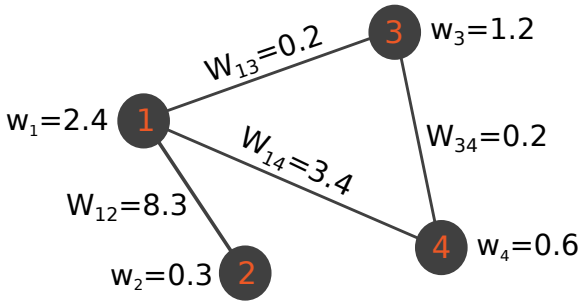

Figure 2.1.: Example of a node- and link-weighted network G with $N_x = 4$ nodes and $M = 4$ links as well. W_{ij} and w_i denote the weight of each node and link, respectively.

between them [Costa et al., 2007, Newman, 2003]. An example of such a network G with a set V of $N_x = 4$ nodes and a set E of $M = 4$ links is sketched in Fig. 2.1.

The topological structure of a network G is usually best represented by its adjacency matrix $\mathbf{A}$ with entries $A_{ij} = 1$ if two nodes i and j are connected via a link and $A_{ij} = 0$ otherwise. In the course of this work, every network is assumed to be *undirected*, which means that no link indicates any form of direction of the interrelations between nodes. Thus, the adjacency matrix $\mathbf{A} = \mathbf{A^T}$ is symmetric and $A_{ij} = A_{ji}$. Also, if not otherwise stated all networks are considered to be free of self-loops, meaning that no node is connected with itself and, thus, $A_{ii} = 0 \; \forall \; i = 1, \ldots, N_x$. For illustration, the adjacency matrix of the undirected example network outlined in Fig. 2.1 reads

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}. \tag{2.1}$$

A more general case, which is not considered in this work, are *directed* networks where links between nodes i and j possess a direction, such that a link specifically starts at a node i and ends at a node j [Costa et al., 2007, Newman et al., 2002]. Consequently, the corresponding adjacency matrix is then asymmetric, $\mathbf{A} \neq \mathbf{A^T}$.

A variety of microscopic and macroscopic metrics have been defined in order to quantify the topological structure of a given network G [e.g., Boccaletti et al., 2006]. They are usually based on specific algebraic operations on the adjacency matrix $\mathbf{A}$ and give insights on the role of single nodes i in the network or they allow for a quantitative intercomparison of a given system with other complex networks. All specific metrics that are relevant for this work are presented in the following sections.

In addition to the observation of binary neighborhood structures, each link between nodes i and j might carry an individual weight W_{ij} that indicates the strength of the

connection between the nodes. Such weights could for example denote the number of flights scheduled between two airports [Bagler, 2008], or the degree of similarity between dynamics at two different locations in the climate system [Donner et al., 2017, Wiedermann et al., 2016c]. Under the presence of such weighted links, the network is consequently represented by its weighted adjacency matrix (or weight matrix) $\mathbf{W}$ with corresponding entries $W_{ij} > 0$ if nodes i and j are connected via a weighted link and $W_{ij} = 0$ otherwise. As an example, the weight matrix of the example network in Fig. 2.1 reads

$$\mathbf{W} = \begin{pmatrix} 0 & 8.3 & 0.2 & 3.4 \\ 8.3 & 0 & 0 & 0 \\ 0.2 & 0 & 0 & 0.2 \\ 3.4 & 0 & 0.2 & 0 \end{pmatrix}. \tag{2.2}$$

In order to quantify the topology of a link-weighted network it is in most cases possible to define corresponding versions of formerly unweighted network characteristics by replacing the adjacency matrix $\mathbf{A}$ with the weight matrix $\mathbf{W}$ in the definition of the associated metrics.

Ultimately, and in addition to weights put on the links, nodes i may also carry individual weights w_i that represent their specific share on the domain of interest under study [Heitzig et al., 2012, Wiedermann et al., 2013]. Thus, node weights could represent the size of an airport or a corresponding city in airline networks or the gross domestic product (GDP) of an economic entity, such as a country, in trade networks [Heitzig et al., 2012, Maluck and Donner, 2015]. The total weight $\mathcal{W}$ of all nodes in the network is computed as the sum over all individual node weights w_i,

$$\mathcal{W} = \sum_{i \in V} w_i. \tag{2.3}$$

As for the link-weighted network characteristics, it is possible to derive node-weighted network metrics from their corresponding unweighted counterparts. However, this procedure is not as straightforward as for the case of a link-weighted case (see above). In particular, Heitzig et al. [2012] introduced a framework to consistently convert standard un-weighted network characteristics into node-weighted counterparts. These so-called *node-splitting-invariant* (n.s.i.) metrics are obtained by implementing the following axiomatic construction scheme:

(a) Sum up weights w_i whenever the unweighted measure counts nodes.

(b) Treat every node $i \in V$ as connected with itself

(c) Allow equality for i and j wherever the original measure involves a sum over distinct nodes i and j.

(d) "Plug in" n.s.i. versions of measures wherever they are used in the definition of other measures.

In order to particularly implement step (b) of the above scheme one has to define the so-called extended adjacency and weight matrices $\mathbf{A}^+$ and $\mathbf{W}^+$, respectively. These matrices specifically incorporate the presence of self-loops and are consequently characterized by including entries on their main diagonal such that $A_{ii} = 1 \; \forall \; i = 1 \ldots, N_x$ and $\mathbf{W}_{ii} > 0 \; \forall \; i = 1 \ldots, N_x$. For the case of a link-weighted graph, the entries W_{ii} must be defined appropriately, for example as the degree of self-similarity of a node i in functional networks [Zemp et al., 2014]. The inclusion of self-loops and redefining the adjacency and weight matrices ensures that network characteristics are unaffected by the removal or the addition of nodes as long as the total weight $\mathcal{W}$ is preserved. The validity of the axioms above can be tested by splitting one node i into two mutually connected nodes i' and i'' with the same neighboring nodes $A_{i'j} = A_{i''j} \; \forall \; j \in V$ and $W_{i'j} = W_{i''j} \; \forall \; j \in V$ such that their combined weights equal that of the former node i, $w_{i'} + w_{i''} = w_i$. In this case, it is possible to show for each network characteristic that its values are unaffected by an appropriate coarse graining or refinement of the network structure. Thus, the measure is invariant under the splitting or merging of single nodes i and j. The thus obtained node-weighted n.s.i. network characteristics are specifically useful to study networks where nodes are sampled inhomogeneously along the domain of interest. One prominent example is the study of functional global climate networks, where due to the spherical geometry of the Earth and the nature of the underlying data sets the density of nodes usually varies strongly between the Equator and the Poles [Heitzig et al., 2012, Wiedermann et al., 2016a,c, Zemp et al., 2014].

Based on the above preliminaries, the following sections introduce a variety of metrics that are utilized to quantify the topological structure of a given network G by means of operations on its adjacency or weight matrix $\mathbf{A}$ and $\mathbf{W}$, respectively. It is noted that beyond the metrics presented here, there exists a large variety of further network characteristics that can be employed to study the micro- and macroscopic structure and function of complex networks [see, e.g., Albert and Barabási, 2002, Boccaletti et al., 2014, Newman, 2003]

2.2. Local (per-node) network characteristics

Local, or per-node, characteristics are computed for each node i in the network individually. They usually give a notion of the nodes' individual importance or centrality in the network and, thus, allow for a comparison of such properties between them. A discrimination between nodes based on such metrics is usually performed when investigating the topology of a single network [e.g., Klimm et al., 2014].

Node degree and strength

As the probably most straightforward network characteristic the degree k_i measures the total number of all nodes in the network that a certain node i is connected to,

$$k_i = \sum_{j \in V} A_{ij}. \tag{2.4}$$

It is computed for each node individually and for the case of an undirected network with no self-loops takes integer values between 0 for an isolated node and $N_x - 1$ for a node that is connected with all other nodes in the network. k_i gives a first impression of a node's individual importance for the function of the entire network under study. Examples include airline networks, where nodes with a high degree k_i are denoted as transfer hubs [Bagler, 2008, Gastner and Newman, 2006] or again functional climate networks where groups of spatially close nodes with a high degree usually indicate large-scale patterns of dominant climate variability [Donges et al., 2011a, Tsonis and Roebber, 2004, Wiedermann et al., 2016c]. The node degrees k_i of the four nodes presented in the example network of Fig. 2.1 read $k_1 = 3$, $k_2 = 1$ and $k_3 = k_4 = 2$.

For the case of a link-weighted network the node degree k_i is transformed into its link-weighted counterpart s_i which measures the total weight of all links connected with a node i,

$$s_i = \sum_{j \in V} W_{ij}. \tag{2.5}$$

s_i is then referred to as the strength of node i. It indicates, e.g., the total trade of a country in economic networks if links are weighted according to bilateral trade volumes [Maluck and Donner, 2015]. For the particular case discussed in this work, s_i is utilized to compute the number of nodes in a climate network that are neighbors of a given node i, weighted according to the cross-correlation of a climate observable measured at both locations. This way, a large number of weak connections between nodes is considered in a qualitatively similar fashion as nodes that display only a few strong ties with their neighbors and, thus, both cases are considered rather equally. Again, for the nodes in the example network (Fig. 2.1) the node strengths are given as $s_1 = 11.9$, $s_2 = 8.3$, $s_3 = 0.4$ and $s_4 = 3.6$.

As discussed above, specifically functional climate networks often require the inclusion of additional node weights to account for an inhomogeneous spatial sampling of the corresponding nodes. In this case, the strength s_i can be converted into a link- and node weighted metric by incorporating steps (a) and (b) of the axiomatic construction scheme presented above [Wiedermann et al., 2016c]. The corresponding

n.s.i. node-strength then reads

$$s_i^* = \sum_{j \in V} w_j W_{ij}^+. \tag{2.6}$$

It again measures the total weight of links that is connected with a node i. However, in contrast to s, s_i^* weighs nodes with larger values of w_j more heavily and, thus, accounts for inhomogeneities in the nodes' particular properties or positions. Assuming unit link-weights $W_{ii} = 1 \; \forall \; i = 1, 2, 3, 4$ in the extended weight matrix $\mathbf{W}^+$ the n.s.i. node-strength of nodes in the example network in Fig. 2.1 reads $s_1^* = 7.17$, $s_2^* = 20.22$, $s_3^* = 1.8$ and $s_4^* = 9$.

Local clustering coefficient

As a second prominent per-node characteristic, the local clustering coefficient $\mathcal{C}_i$ measures the fraction of fully connected triples of nodes that are centered at a node i,

$$\mathcal{C}_i = \frac{1}{k_i(k_i - 1)} \sum_{j \neq l \in V} A_{ij} A_{jl} A_{li} \in [0, 1]. \tag{2.7}$$

High values of $\mathcal{C}_i$ indicate nodes that tend to connect with other nodes that are themselves mutually connected. In this case, the corresponding nodes tend to form strongly connected components within the network that are commonly referred to as communities [Fortunato, 2010, Tsonis et al., 2010]. Such communities could be composed of, e.g., authors from the same scientific field in citation networks [Newman, 2001], or again areas of statistically similar dynamics in climate networks [Tsonis and Roebber, 2004, Tsonis et al., 2010]. Note that for nodes with degree $k_i \leq 1$, one defines $\mathcal{C}_i = 0$ as no triple can be formed by the node i [Costa et al., 2007].

As an analogue to the node degree, one may define link weighted versions of the local clustering coefficient as well. However, for this metric such a conversion is not necessarily unique and a variety of weighted versions of $\mathcal{C}_i$ have been defined [Saramäki et al., 2007] that mostly differ with respect to their normalization. As in this work the local clustering coefficient is only applied to unweighted real-world networks, no further link- or node-weighted refinement of this measure is necessary.

Ultimately, it remains to note that a large variety of further local statistics to quantify the topological role of certain nodes i in the network have been defined in the past. For a comprehensive summary of such metrics see, e.g, Albert and Barabási [2002], Boccaletti et al. [2006], Newman [2003].

2.3. Global (macroscopic) network characteristics

In addition to the local per-node characteristics, global network measures quantify the macroscopic structure of a complex network by a single characteristic number. Their computation is useful to intercompare the topology of a given complex network with that of other systems or models under study. This works relies (depending on the specific application) on commonly used global measures that are used when discussing a network's macroscopic structure, specifically with respect to categorizing networks into different classes, such as the attribution of the small world property [Milgram, 1967, Watts and Strogatz, 1998], which is characteristic for many real-world complex networks as it implies a low degree of separation between different entities, e.g., nodes in the underlying system.

Link density

As one of the most fundamental macroscopic network characteristics the link density d gives the ratio between the actual number of links M in the network and the maximum possible number of links,

$$d = \frac{2M}{N_x(N_x-1)} = \frac{1}{N_x(N_x-1)} \sum_{i \neq j \in V} A_{ij} \in [0,1]. \tag{2.8}$$

High values of d indicate a strong connectivity of the network. In particular for the case of functional (climate) networks, the link density d is intimately connected with the global threshold T that is applied to construct the network from a given similarity matrix $\mathbf{S}$ [Donges et al., 2009b]. Here, the link density is often fixed to small values $d \leq 0.05$ and the corresponding threshold T is computed as the respective quantile of $\mathbf{S}$ (see Chap. 3 for details).

Global clustering coefficient and average path length

The global clustering coefficient $\mathcal{C}$ gives the probability to find connected triples, i.e., closed triangles formed by links in the network adjacent to a randomly selected node [Watts and Strogatz, 1998]. It is defined as the arithmetic mean taken over all local clustering coefficients,

$$\mathcal{C} = \frac{1}{N_x} \sum_{i \in V} \mathcal{C}_i \tag{2.9}$$

with $\mathcal{C}_i$ as given in Eq. (2.7). $\mathcal{C}$ gives a notion of the entire network's tendency to form clusters of closely connected triples of nodes and relates to the corresponding chance of finding distinct communities [Newman, 2003]. Specifically, high values of $\mathcal{C}$ thus indicate the presence of one or more closely connected clusters of nodes while

low values indicate a rather homogeneous distribution of links along the network and the absence of a distinct community structure [Newman, 2010].

Accompanying the global clustering coefficient $\mathcal{C}$, the average path length $\mathcal{L}$ gives the average number of links along shortest paths between two randomly chosen nodes. Given that $\mathcal{L}_{ij}$ denotes the number of such steps between two nodes i and j, the average path length follows as

$$\mathcal{L} = \frac{1}{N_x(N_x-1)} \sum_{i \neq j \in V} \mathcal{L}_{ij}. \tag{2.10}$$

In the case when there exists no path between i and j one sets $\mathcal{L}_{ij} = N_x - 1$ [Albert and Barabási, 2002]. $\mathcal{L}$ thus estimates the average topological distance between two nodes or represented entities thereof and is closely related to the notion of *degrees of separation* in human societies [Milgram, 1967]. A combined assessment of $\mathcal{C}$ and $\mathcal{L}$ allows for a quantification of such degrees of separation and the corresponding assignment of the small-world property to a complex network [Davidsen et al., 2002, Watts and Strogatz, 1998]. This property is particularly characterized by high values of $\mathcal{C}$ indicating a distinct community structure and low values of $\mathcal{L}$ indicating a low degree of separation. It has been observed in a variety of studies that the small-world characteristic holds for many real-world complex systems and networks, such as communication [Newman et al., 2002], transportation [Guimerà et al., 2005], or social networks [Davidsen et al., 2002, Ebel et al., 2002, Korte and Milgram, 1970].

Transitivity

Notably, $\mathcal{C}$ gives equal weight to the contributions of all nodes independent of their respective degree, which can lead to problematic properties of this measure when being evaluated for very heterogeneous networks with many nodes of degree $k_i \leq 1$ which consequently have zero local clustering coefficients. This issue is usually not very prominent for real-world networks, where isolated nodes are often excluded before assessing the network's topological characteristics. However, in the context of climate networks it might well be that a large number of such isolated nodes appears due to large dynamical dissimilarities in the climatic observations. Hence, in the latter context, it can be useful to study a closely related measure, the network transitivity $\mathcal{T}$, as an alternative to the global clustering coefficient $\mathcal{C}$. It is defined as

$$\mathcal{T} = \frac{\sum_{i \neq j \neq l \in V} A_{ij} A_{jl} A_{li}}{\sum_{i \neq j \neq l \in V} A_{ij} A_{jl}} \in [0, 1]. \tag{2.11}$$

$\mathcal{T}$ is based on an alternative measure of quantifying clustering in a complex network and a formal description of the following widely used definition [Newman, 2003,

2010] where

$$\text{Clustering} = \frac{3 \times \text{number of triangles}}{\text{number of loose triple}}. \tag{2.12}$$

Note that, in contrast to $\mathcal{C}$, $\mathcal{T}$ is defined without the inclusion of a corresponding local counterpart. However, propositions regarding the derivation of such a node-wise transitivity have been put forward by Radebach et al. [2013].

In order to apply $\mathcal{T}$ specifically in the climate network context it is again useful to define a link-weighted correspondent in order to weigh links according to the statistical interdependence between two nodes. Among other definitions [Newman, 2001, Saramäki et al., 2007] one possible way of defining a link-weighted correspondent to the unweighted transitivity is by consequently replacing the adjacency matrix $\mathbf{A}$ by the weight matrix $\mathbf{W}$ as proposed by Zhang and Horvath [2005] and Kalna and Higham [2006]. This yields the weighted transitivity as

$$\mathcal{T} = \frac{\sum_{i \neq j \neq l \in V} W_{ij} W_{jl} W_{li}}{\sum_{i \neq j \neq l \in V} W_{ij} W_{il}} \tag{2.13}$$

This metric is not necessarily bounded by one, but only as long as $\max_{ij} W_{ij} \leq 1$. For cases where this property is not fulfilled it is advisable to divide all weights W_{ij} by the maximum possible weight $\max_{ij} W_{ij}$ [Saramäki et al., 2007]. For the remainder of this work the specific symbol $\mathcal{T}$ always refers to the link-weighted version of transitivity.

As an analogue to the definition of the n.s.i. node-strength s^* a node- and link-weighted version of network transitivity can be derived by applying the construction scheme in Sec. 2.1 to Eq. (2.13). This yields the n.s.i. weighted transitivity $\mathcal{T}^*$ as [Wiedermann et al., 2016c]

$$\mathcal{T}^* = \frac{\sum_{i,j,l \in V} w_i W^+_{ij} w_j W^+_{jl} w_l W^+_{li}}{\sum_{i,j,l \in V} w_i W^+_{ij} w_j W^+_{il} w_l} \tag{2.14}$$

Hence, $\mathcal{T}^*$ gives the link- and node-weighted fraction of completely linked triples of nodes and measures how strongly nodes in the network tend to form closely connected clusters. Likewise the global and local clustering coefficient, it naturally takes low values under the absence of such tendencies and high values if a distinct community structure and a corresponding strong localization of higher local link-densities is present. Particularly, in a purely random network where each link between nodes is present with the same probability, $\mathcal{T}$ and $\mathcal{T}^*$ would naturally take very low values, i.e., on average equal to the link density in the standard case of no specific link and node-weights [Erdös and Rényi, 1960].

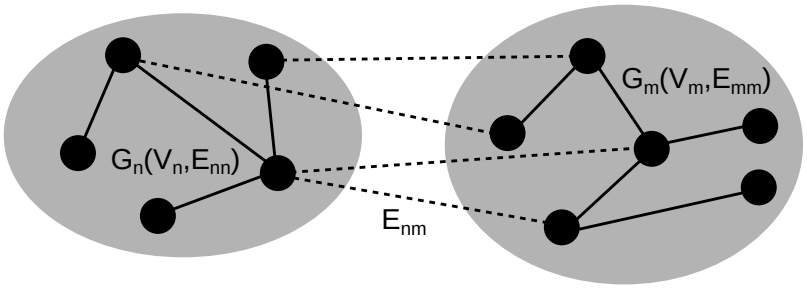

Figure 2.2.: A set of two subnetworks $G_n = (V_n, E_{nn})$ and $G_m = (V_m, E_{mm})$ connected via both internal (solid lines) and cross-links (dashed lines).

Hamming distance

In order to derive a final macroscopic network characteristic used in this work, consider the case of two undirected networks $G = (V, E)$ and $G' = (V, E')$ with a common set of nodes and the same number of links $M = |E| = |E'|$ represented by adjacency matrices $\mathbf{A}$ and $\mathbf{A}'$. The Hamming distance $\mathcal{H}$ then provides a measure of the dissimilarity between the two sets of links E and E' [Hamming, 1950, Radebach et al., 2013],

$$\mathcal{H} = \frac{1}{4M} \sum_{i,j \in V} |A'_{ij} - A_{ij}| \in [0, 1]. \tag{2.15}$$

A Hamming distance of $\mathcal{H} = 0$ implies that the two networks G and G' have identical sets of links ($E = E'$), while $\mathcal{H} = 1$ indicates that the two sets of links have entirely dissimilar entries ($E \cap E' = \emptyset$). In the scope of this work, $\mathcal{H}$ is utilized to assess the dissimilarity between two given networks under study. In particular it will by employed in the context of spatial networks (Chap. 7) to estimate the dissimilarity between a given network and a randomized counterpart.

2.4. Generalization to coupled networks

In addition to characteristics for networks composed of a single layer or subnetwork (as the ones introduced above), Donges et al. [2011b] proposed a framework to also quantify interactions between a set of subnetworks that form a larger system of so-called *interacting networks* or *coupled networks*. In this case, a network $G(V, E)$ is composed of a set of N_G subnetworks $G_n = (V_n, E_{nn}), n = 1, \ldots, N_G$ with sets of nodes V_n such that each node belongs to exactly one subnetwork (i.e. $V = \cup_n V_n$ and $V_n \cap V_m = \emptyset \ \forall \ n, m = 1, \ldots, N_G; \ n \neq m$). N_{xn} then gives the number of nodes in subnetwork G_n. Likewise, the set of links E splits into internal link sets E_{nn} (connecting nodes within a subnetwork) and cross-link sets E_{nm} connecting nodes $i \in V_n$ with nodes $j \in V_m$ in the subnetworks G_n and G_m, respectively [Donges

et al., 2011b]. An example of such an interacting network that is composed of two subnetworks is given in Fig. 2.2. The adjacency matrix $\mathbf{A}$ of the entire network G then decomposes into *internal* adjacency matrices $\mathbf{A}_n$, representing the internal link structure within one subnetwork G_n, and *cross*-adjacency matrices $\mathbf{A}_{nm}$, representing links between nodes $i \in V_n$ and $j \in V_m$. Thus,

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_{1,2} & \cdots & \mathbf{A}_{1,N_G} \\ \mathbf{A}_{2,1} & \mathbf{A}_2 & \cdots & \mathbf{A}_{2,N_G} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{N_G,1} & \mathbf{A}_{N_G,2} & \cdots & \mathbf{A}_{N_G} \end{pmatrix}. \tag{2.16}$$

As an analogue to single layer networks, the cross-adjacency matrices are symmetric for the case of an undirected network, such that $\mathbf{A}_{nm} = (\mathbf{A}^{\mathbf{T}})_{mn}$.

In order to account for the different density of links between and within subnetworks the notion of the link density d is refined such that the link densities within each subnetwork and between two subnetworks are considered separately. In particular, the *internal* link density d_n within each subnetwork G_n is computed in full analogy to Eq. (2.8) as

$$d_n = \frac{2M_n}{N_{xn}(N_{xn}-1)} = \frac{1}{N_{xn}(N_{xn}-1)} \sum_{i \neq j \in V_n} A_{ij} \in [0,1], \tag{2.17}$$

where M_n and N_{xn} denotes the number of links and nodes within subnetwork G_n, respectively. In addition, the *cross*-link density d_{nm} measuring the normalized number of links between two subnetworks G_n and G_m is given as

$$d_{nm} = \frac{2M_{nm}}{N_{xn}N_{xm}} = \frac{1}{N_{xn}N_{xm}} \sum_{\substack{i \in V_n \\ j \in V_m}} A_{ij} \in [0,1]. \tag{2.18}$$

Here, M_{nm} denotes the number of cross-links connecting nodes i in subnetwork G_n with nodes j in subnetwork G_m.

Note that as for now there exists no standard terminology for the assessment of such interacting networks. In the case that the sets of nodes are identical for all subnetworks the resulting networks are often referred to as so-called multilayer networks [Kivelä et al., 2014]. Also there is ongoing debate on whether such networks are best represented by a *super-adjacency matrix* similar to the definition given in Eq. (2.16) or if a tensor-based notation is more suitable. Since the aim of the present study is to apply this formalism to the assessment and discrimination of different layers in coupled interacting climate networks, this work relies on the notation and definitions by Donges et al. [2011b] as they are introduced above.

Local measures

Interacting networks can be assessed in full analogy to the single-layer counterparts as presented in Sec. 2.2 and 2.3. For quantifying local cross-network interactions in such coupled networks this work relies on two measures, n.s.i. cross-degree k_i^{m*} and n.s.i. local cross-clustering coefficient $\mathcal{C}_i^{m*}$, that were introduced by Wiedermann et al. [2013] and (for the case of the n.s.i. cross-degree) by Feng et al. [2012]. They are derived from their unweighted counterparts defined by Donges et al. [2011b] in the same way as the single layer measures introduced above, i.e., by applying the four-step construction mechanism introduced by Heitzig et al. [2012] (Sec. 2.1).

The two measures, k_i^{m*} and $\mathcal{C}_i^{m*}$ are defined to capture the topological structure of interactions between two disjoint subnetworks, i.e., nodes $i \in V_n$ and $j \in V_m$. They read

$$k_i^{m*} = \sum_{j \in V_m} w_j A_{ij}^+, \tag{2.19}$$

$$\mathcal{C}_i^{m*} = \frac{1}{\left(k_i^{m*}\right)^2} \sum_{j,l \in V_m} A_{ij}^+ A_{jl}^+ A_{li}^+ w_j w_l \in [0,1]. \tag{2.20}$$

Again, $w_\bullet$ denotes the weights of individual nodes and $A_{\bullet\bullet}^+$ denotes entries of the extended (super-) adjacency matrix of the entire network G that is composed of all subnetworks G_n. In contrast to the unweighted cross-degree

$$k_i^m = \sum_{j \in V_m} A_{ij} \tag{2.21}$$

which simply counts nodes $j \in V_m$ that are connected with $i \in V_n$, k_i^{m*} is proportional to the share of the total weight $\mathcal{W}_m$ of all nodes $j \in V_m$ that $i \in V_n$ is connected with. Again, these node-weighted metrics are in the course of this work particularly useful when investigating functional climate networks with inhomogeneously sampled nodes.

Similar to k_i^{m*}, $\mathcal{C}_i^{m*}$ no longer relies on the counting of distinct fully connected node triples in the network (as, e.g., the classical local clustering coefficient in Eq. (2.7) [Newman, 2003]) but on the weighted sum of occurrences of triples of connected shares within the two subnetworks. It gives the probability that a share of $\mathcal{W}_n$ represented by a node $i \in V_n$ is connected with two mutually connected shares of $\mathcal{W}_m$ areas in the opposite subnetwork. A local accumulation of such connected triples again represents clusters of closely connected nodes. With respect to the global climate system, $\mathcal{C}_i^{m*}$ thus estimates how likely areas in two different climatic fields or subsystems form clusters of statistical equivalence between them.

In order to make the n.s.i. cross-degree k_i^{m*} comparable between different subnetworks, it is normalized by the total possible weight that nodes $i \in V_n$ can be

connected with,

$$\kappa_i^{m*} = \frac{\sum_{j \in V_m} w_j A_{ij}^+}{\mathcal{W}_m} \in [0, 1]. \tag{2.22}$$

In the spirit of earlier work by Donges et al. [2012] and Donner et al. [2010], this quantity is referred to as the *n.s.i. cross-degree density*. For the case of a single-layer network, a measure similar to the n.s.i. cross-degree density has been introduced by Tsonis et al. [2006] in terms of the *area weighted connectivity*, which quantifies the share on the subdomain of interest represented by the entire network G that is connected with any nodes $i \in V$.

Generally, Wiedermann et al. [2013] and Zemp et al. [2014] showed that the weighted local cross-network measures improve the representation of a network's topology with inhomogeneous node density within the domain of interest in comparison with its unweighted counterparts.

Global measures

In addition to local (per node) network measures the present work also aims to characterize the macroscopic interaction structure of each pair of interacting networks by means of global network properties. For interacting networks a variety of unweighted measures have been proposed by Donges et al. [2011b]. Here, the construction mechanism by Heitzig et al. [2012] (Sec. 2.1) is utilized to convert two of them into their weighted counterparts as well.

N.s.i. global cross-clustering coefficient

The global cross-clustering coefficient $\mathcal{C}_{nm}$ of a subnetwork G_n gives the probability that for a randomly chosen node $i \in V_n$ one finds neighbors $j, l \in V_m$ that are mutually linked. It is defined as the arithmetic mean of all local cross-clustering coefficients $\mathcal{C}_i^m$,

$$\mathcal{C}_{nm} = \frac{1}{N_{xn}} \sum_{i \in V_n} \mathcal{C}_i^m. \tag{2.23}$$

This measure can be converted into its n.s.i. counterpart by calculating the weighted mean of all values of $\mathcal{C}_i^{m*}$ that is given in Eq. (2.20),

$$\mathcal{C}_{nm}^* = \frac{1}{\mathcal{W}_n} \sum_{i \in V_n} w_i \mathcal{C}_i^{m*}. \tag{2.24}$$

Again, analogously to the interpretation of the local n.s.i. measures, $\mathcal{C}_{nm}^*$ no longer only measures pure node-wise triangular structures but takes into account the share

of total weight involved in their formation. Generally, large values of $\mathcal{C}^*_{nm}$ (which are induced by a dominance of connected triples between the two subnetworks under consideration) indicate strong transitivity in the underlying network structure.

N.s.i. cross-transitivity

Similar to $\mathcal{C}_{nm}$, the cross-transitivity $\mathcal{T}_{nm}$ gives the probability that two randomly drawn nodes $j, l \in V_m$ are connected if they have a common neighbor $i \in V_n$. It is given as

$$\mathcal{T}_{nm} = \frac{\sum_{i\in V_n}\sum_{j\neq l\in V_m} A_{ij}A_{jl}A_{li}}{\sum_{i\in V_n}\sum_{j\neq l\in V_m} A_{ij}A_{il}}. \tag{2.25}$$

Like $\mathcal{C}_{nm}$, the cross-transitivity is a measure of organization with respect to the cross-link structure in a coupled network [Donges et al., 2011b]. However, in contrast to $\mathcal{C}_{nm}$, $\mathcal{T}_{nm}$ takes into account the increasing influence of nodes with high cross-degree and weighs them more heavily than nodes with low cross-degree. More importantly it ignores nodes with no links into the opposite layer, since these nodes display a zero cross-degree. The node-weighted variant of $\mathcal{T}_{nm}$ can be written as

$$\mathcal{T}^*_{nm} = \frac{\sum_{i\in V_n}\sum_{j,l\in V_m} w_i A^+_{ij} w_j A^+_{jl} w_l A^+_{li}}{\sum_{i\in V_n}\sum_{j,l\in V_m} w_i A^+_{ij} w_j w_l A^+_{jl}} = \frac{\sum_{i\in V_n} w_i (k^{m*}_i)^2 \mathcal{C}^{m*}_i}{\sum_{i\in V_n} w_i (k^{m*}_i)^2}. \tag{2.26}$$

Note, that both $\mathcal{C}^*_{nm}$ and $\mathcal{T}^*_{nm}$ similarly evaluate the tendency of nodes to form clusters of closely connected triples between two subnetworks G_n and G_m. They are derived, however, in a disjoint manner. One measure, $\mathcal{C}^*_{nm}$ is computed as the weighted average taken over $\mathcal{C}^{m*}_i$. In contrast, despite suggestions by Radebach et al. [2013] to decompose the global transitivity into local contributions, the n.s.i. cross-transitivity $\mathcal{T}^*_{nm}$ is defined solely as a global network measure with no direct local counterpart. It is further important to note that n.s.i. cross-transitivity and n.s.i. global cross-clustering coefficient are commonly asymmetric in the sense that $\mathcal{T}^*_{nm} \neq \mathcal{T}^*_{mn}$ and $\mathcal{C}^*_{nm} \neq \mathcal{C}^*_{mn}$.

This section concludes the preliminary observations which form the basis of all following investigations. The following first part of this work discusses the application of complex network theory as a framework to discriminate different types of dynamics in the climate system. It heavily relies on the definitions of node- and (in parts) link-weighted network characteristics which have been introduced in the course of this chapter. After the presentation and discussion of the two climate applications, this work moves towards the study of structural (real-world) networks and discusses possible biases of specifically the global clustering coefficient and average path length with respect to geographical constraints that are present in a variety of complex

networks. Ultimately, this work proposes, as a compliment to the characteristics presented in this work, an information theoretic approach in order to quantitatively assess a network's actual *complexity* and further discriminate different classes and types of networks based on such a complexity measure.

Part I.

Functional (climate) networks

Chapter 3.

Climate network construction

The previous chapter introduced as a preliminary the general framework and underlying theory that is related to the study of complex networks. Specifically, the notion of an adjacency ($\mathbf{A}$) or weight matrix ($\mathbf{W}$) as well as the concept of individual node weights $w_\bullet$ was introduced. The following first major part of this work is devoted to the application of complex network theory to study functional or statistical interdependencies between dynamics that are observed at different nodes. The resulting network representation of such a system is usually denoted as a so-called *functional network* [Bullmore and Sporns, 2009, Castillo, 1998].

In this work the specific applications are chosen from a climatic context where functional network theory has given rise to a comparatively new field of research that is commonly referred to as *climate network analysis* [Donges et al., 2009b, Donner et al., 2017, Tsonis et al., 2006]. In climate networks, nodes represent geographical locations on or parallel to the Earth's surface where one or more time series of climatic observables are available. Links are then put between nodes according to some statistical interdependency between the time series at different locations. Before discussing this work's specific applications of climate networks to study and discriminate (i) different phases of the El Niño Southern Oscillation (Chap. 4) and (ii) different characteristics of northern hemispheric Ocean-Atmosphere interactions (Chap. 6), this chapter focuses on the necessary steps to construct a complex network as an appropriate representation of interdependencies within such climatic systems.

The introductions given here are kept as general as possible in order to show the applicability of functional networks to study interactions and interdependencies within a variety of complex systems, such as the brain [Wang et al., 2009] or trade relations [Maluck and Donner, 2015]. However, wherever necessary the presented concepts will be discussed with respect to the specific application of studying spatially extended climate data. In particular, three consecutive steps are discussed: (i) The preprocessing of input data (Sec. 3.1), (ii) the estimation of functional or statistical dependencies between different entities in the input data (Sec. 3.2) and (iii) the construction of the resulting network's adjacency matrix $\mathbf{A}$ or weight matrix $\mathbf{W}$ (Sec. 3.3) which can then be analyzed with the tools and metrics that were presented

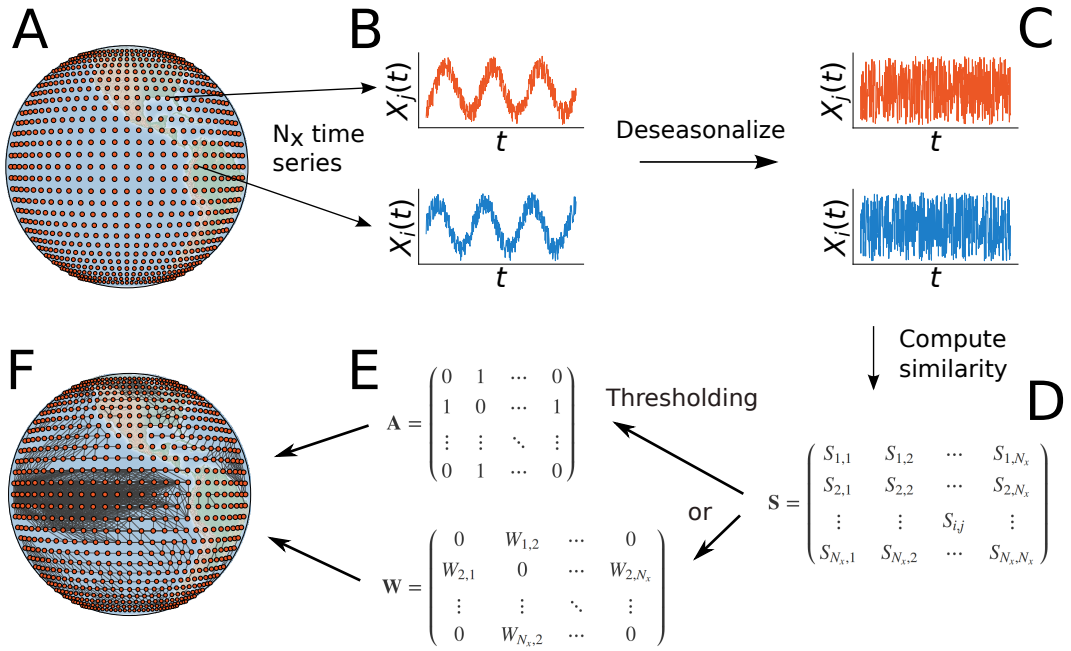

Figure 3.1.: Schematic illustration of the construction of a climate network from a global gridded climatic data set. In this case, grid points (which are identified as corresponding nodes in the climate network) are located at regularly spaced longitudinal and latitudinal positions (A) and each point provides climatic observations in terms of one or more time series (B). After an appropriate preprocessing, i.e., performing a grid-point wise seasonal detrending (C), the pairwise similarity between dynamics at all grid points is assessed (D) and a climate network is obtained by thresholding the corresponding similarity matrix (E). From this, the final climate network representation is obtained (F). (adapted from Donner et al. [2017])

in Chap. 2. Ultimately, as analogue to the introduction of interacting networks as a complement to the study of single-layer networks, the above concepts are generalized to also study *coupled climate networks* that are specifically useful to quantify statistical interdependencies between different climatic observables (Sec. 3.4) [Donges et al., 2011b, Wiedermann et al., 2016a]

This chapter is in parts based on methodological considerations that were also put forward in P6 [Wiedermann et al., 2016c], P8 [Wiedermann et al., 2016a] and P9 [Donner et al., 2017].

3.1. Data acquisition, preprocessing and node weights

Functional networks are usually constructed from a set of N_x discrete time series $X_i(t),\ t = 1, 2, \ldots, N_t$ where N_t gives their respective number of temporal sampling points [Castillo, 1998, Donges et al., 2009a, Wang et al., 2009]. Nodes i in the network

then represent each of the time series $X_i(t)$ and links between those nodes indicate some functional or statistical dependence between them. In climate networks, nodes represent for example grid points along the Earth's surface (Fig. 3.1A), each of which represents dynamics at their specific location that are provided in terms of a climatic time series (Fig. 3.1B)

Generally, each time series represents a part of the dynamics of a certain system under study, such as EEG signals representing dynamics of the brain in neural networks [Smit et al., 2008], or records of a specific climate variable in climate networks [Tsonis and Roebber, 2004]. The steps required to prepare such data in order to perform a meaningful analysis vary strongly with the kind of application.

In climate networks, it must be ensured that the individual time series $X_i(t)$ represent a comparable amount of information on climate variability at the considered time scale, i.e, they should be free of any known underlying periodicities and trivial dominant modes [Donner et al., 2017]. Such fundamental periodicities could be the annual solar cycle which induces a strong seasonality in most observables, or oscillations on longer time scales, such as the Atlantic Multidecadal Oscillation [Dima and Lohmann, 2007]. As this work primarily focuses on interactions at the daily to monthly temporal scale, the annual cycle must be removed from all input signals prior to any further analysis. Otherwise, any estimation of similarity or correlation between time series will essentially be dominated by their similarity at the annual scale, while evolutions at daily to weekly scales will be masked by the annual variability component. Thus, for a data set with a yearly number of u temporal sampling points ($u = 365$ for daily data and $u = 12$ for monthly data) each time series $X_i(t)$ is transformed to a so-called *anomaly* time series in the following way,

$$X_i(t) \to X_i(t) - \langle X_i(t) \rangle_{t \in t'}. \tag{3.1}$$

Here, t' is the set of temporal sampling points that represent the same time during each year in the series, $t' \in \{t \mod u, u+(t \mod u)), 2u+(t \mod u)), \ldots, t, \ldots, N_t - u + (t \mod u)\}$. Note that this procedure is not unique to climate network analysis, but common practice in many techniques of statistical climatology [e.g., Feng and Dijkstra, 2014, Handorf and Dethloff, 2012, Larkin and Harrison, 2005]. Having performed an appropriate deseasonalization of the data (i.e., by obtaining the above definition of climatic anomalies), the considered data set can be used for the network construction (Fig. 3.1C).

A second requirement relates to the most likely heterogeneous spatial distribution of observations [Heitzig et al., 2012, Wiedermann et al., 2013]. If, for example, data is provided from direct measurements at weather stations, the density of these stations per region is likely to differ along the entire spatial domain that is considered in the study [Rheinwalt et al., 2012]. Similar problems arise when one considers reanalysis or climate model data where data is usually provided at a grid with regular

latitude/longitude spacing [Kalnay et al., 1996]. Here, the density of grid points commonly increases rapidly towards the poles. In both such cases, the inhomogeneous sampling is likely to induce some systematic bias into the construction of the network giving more weight to those areas with higher densities of nodes [Heitzig et al., 2012, Tsonis et al., 2006]. Specifically, in cases where networks are constructed from similarity assessments between dynamics at different stations or grid points this issue becomes crucial as spatially close nodes are more likely to be connected than spatially distant ones. In this spirit, areas with a larger density of nodes tend to also accumulate a larger number of links, thus, causing network characteristics to overestimate the role of these regions in the resulting climate networks [Radebach et al., 2013].

In order to avoid these biases, two approaches have been introduced in the past. One possible way to circumvent spatial inhomogeneities is to interpolate all data to a geodesic grid with homogeneous node distances and thus, homogeneous density, along the entire spatial domain [Radebach et al., 2013]. Another, more common approach is to assign weights to each node according to the area along the considered spatial domain, e.g., the Earth's surface that it represents [Feng et al., 2012, Tsonis and Roebber, 2004, Zemp et al., 2014]. Hence, in regions with a large density of nodes, the respective individually represented areas are low while for regions with a low density, the corresponding represented areas become larger. In case of station data an appropriate node weight could therefore obtained from a Voronoi-Tessalation of the entire region of study according to the position of the individual stations [Rheinwalt et al., 2012]. In case of gridded climate data it is sufficient to weigh each node, i.e., grid point, i according to its latitudinal position λ_i on the grid [Tsonis et al., 2006, Wiedermann et al., 2016a,c],

$$w_i = \cos \lambda_i. \tag{3.2}$$

This way, nodes located densely at the Poles are weighed less than nodes located at the Equator. After the final construction of the climate network representation these weights w_i can be plugged into the estimation of node-weighted network characteristics, such as the n.s.i. node-strength (Eq. (2.6)), that were introduced in Chap. 2.

3.2. Similarity estimation

The second step in the construction of a functional network, and specifically a climate network, is to obtain the $N_x \times N_x$ similarity matrix $\mathbf{S}$ with entries S_{ij} that gives the pairwise statistical or functional interdependency between pairs of nodes i and j (Sec. 3.1C). The choice of the similarity measure certainly depends on the specific application. However, most previous studies on climate networks relied on the pairwise Pearson correlation coefficient at zero lag ρ_{ij} as an appropriate measure

of similarity between two time series $X_i(t)$ and $X_j(t)$ [Donges et al., 2009b, Hlinka et al., 2014, Radebach et al., 2013, Tsonis et al., 2010],

$$\rho_{ij} = \frac{\sum_t (X_i(t) - \overline{X}_i)(X_j(t) - \overline{X}_j)}{\sum_t (X_i(t) - \overline{X}_i)^2 \sum_t (X_j(t) - \overline{X}_j)^2}, \quad i, j = 1, \ldots, N_x. \tag{3.3}$$

Here $\overline{X}_i$ and $\overline{X}_j$ denote the temporal mean values of each time series. Since the correlation coefficient is evaluated at lag zero ρ_{ij} is symmetric in i and j and, hence, $\rho_{ij} = \rho_{ji}$. Consequently the corresponding $N_x \times N_x$ correlation matrix $\mathbf{C}$ with entries ρ_{ij} is symmetric as well, $\mathbf{C} = \mathbf{C^T}$.

With the pairwise correlation matrix $\mathbf{C}$ at hand, it remains to note that potential feedbacks and interdependencies in the climate system can be of positive and negative nature, even within a single climatic observable. An example of such a positive feedback would be an increase in local temperature at location i which leads to increased evapotranspiration, especially from the oceans. This yields an increase in the concentration of water vapor in the atmosphere that enhances the greenhouse effect and thus, yields yet again an increase in temperature at a possibly different location j. Correspondingly the estimated correlation ρ_{ij} between dynamics at location or node i and j are likely to be positive as well. In contrast, as an example for a negative feedback cycle consider again an increase in temperature at location i leading to an increase in evapotranspiration from the oceans. However, this increase in evapotranspiration potentially also enhances cloud formation in the atmosphere which may ultimately lead to a potential cooling at a different location j. This scenario would most likely imply a negative correlation ρ_{ij}.

As climate networks generally aim to quantify any kind of strong statistical interrelationship between observations at different locations it is therefore advisable to account for both, positive and negative feedback loops, simultaneously. Hence, as a final step of similarity estimation, the absolute value of pairwise cross-correlation is considered such that the entries S_{ij} of $\mathbf{S}$ read

$$S_{ij} = |\rho_{ij}|. \tag{3.4}$$

Certainly, this method of similarity estimation is only one among many other possible options. However, most climatic studies beyond climate network research also rely on similar statistical estimators [e.g. Ashok et al., 2007, Graf and Zanchettin, 2012, Handorf and Dethloff, 2012, Hannachi et al., 2007]. Thus, in order to make the results of this work comparable with observations from classical statistical climatology the Pearson correlation coefficient as described above is also employed here to ensure a maximum degree of consistency.

Note that in addition to the study of Pearson correlation at lag zero, previous studies have also incorporated a more general variant of this metric at distinct positive

nonzero lags τ [Gozolchiani et al., 2008, Ludescher et al., 2014]. Since the exact lags between dynamics at different locations are usually unknown, it has been suggested that for such cases the correlation coefficient $\rho_{ij}(\tau)$ is computed for a range of lags τ and the respective maximum value $\max_\tau \rho_{ij}(\tau)$ over all choices of τ then qualifies as a measure of statistical similarity between two nodes i and j, $S_{ij} = \max_\tau \rho_{ij}(\tau)$.

For the sake of completeness, a few other options for estimating similarity in climatic data sets shall be mentioned here as well [Donner et al., 2017]. First, if one aims to assess relationships beyond an assumption of linearity, mutual information has been proven as an appropriate tool that allows for capturing general statistical interdependencies of arbitrary functional form between two variables [Donges et al., 2009a] or corresponding order patterns at different lags [Barreiro et al., 2011, Deza et al., 2013, Tirabassi and Masoller, 2013]. This framework has been successfully applied in order to detect mutual relationships in the climate system at different intra- and inter-annual time scales. In order to study spatio-temporal signatures of extreme precipitation events, event synchronization has been widely applied as a similarity measure to mainly construct regionally constraint climate networks [Boers et al., 2014a, 2013, 2014b, Malik et al., 2010, 2011, Stolbova et al., 2014]. Finally, another notable variant of climate network analysis is based on reconstructing pairwise or non-pairwise causal interrelationships between different processes that are represented by directed links in the network [Runge et al., 2012a, 2013]. The key aspect here is the ability to filter out potentially misleading effects of auto-dependencies, common drivers, or indirect couplings [Ebert-Uphoff and Deng, 2012, Kretschmer et al., 2016, Runge et al., 2012b, 2015].

3.3. Unweighted and weighted network construction

In the last step of functional network construction, the similarity matrix $\mathbf{S}$ with entries S_{ij} is converted into an unweighted or weighted adjacency matrix, $\mathbf{A}$ or $\mathbf{W}$, respectively. Interpreting $\mathbf{S}$ directly as as weight matrix $\mathbf{W}$ may cause several conceptual problems. First, it is not ensured that all entries of $\mathbf{S}$ denote similarity estimations that can be considered significant. For the case of cross-correlation it should, in analogy with most climatic applications, therefore at least be ensured that the considered statistical relations are significant according to a student's t-test [Tsonis et al., 2006]. Secondly, a fully connected network may contain a large portion of trivial information, such that possibly interesting topological features are suppressed by a large portion of weak links between nodes [Hlinka et al., 2014].

Addressing both issues, is has become common practice to set a fixed and global threshold T such that only a certain fraction (i.e. the link density d, Eq. (2.8)) of strongest correlations is represented by links in the resulting climate network [Donges et al., 2009a, Tsonis and Roebber, 2004, Tsonis et al., 2006]. In this case only

similarity estimations between two nodes i and j that are larger than the considered threshold T are represented by links in the resulting network (Fig. 3.1E). Thus, the entries A_{ij} of the adjacency matrix are obtained as

$$A_{ij} = \Theta(S_{ij} - T)(1 - \delta_{ij}). \tag{3.5}$$

Here, δ_{ij} denotes Kronecker's delta with $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise, which ensures that no trivial self-loops are present in the resulting network.

In most applications the threshold T is obtained numerically by computing the quantile π of the entries $S_{\bullet\bullet}$ of $\mathbf{S}$ that corresponds to a desired link density d (Eq. (2.8)). In practice, the link density is usually fixed to comparatively low values $d \leq 0.05$ ensuring that weak and non-significant values of similarity are not considered in the network representation of the observable under study [Donges et al., 2009a, Hlinka et al., 2014, Radebach et al., 2013, Stolbova et al., 2014]. However, it might well be that the actual range of thus included similarity values is considerably large [Wiedermann et al., 2016c]. In this case it proved appropriate to weigh the links in the resulting climate network according to the actual degree of similarity S_{ij} between connected nodes i and j. This then yields a weighted adjacency matrix $\mathbf{W}$ with entries W_{ij} as introduced in Sec. 2.1 which in the case of functional climate networks reads

$$W_{ij} = \Theta(S_{ij} - T)(1 - \delta_{ij})S_{ij} = A_{ij}S_{ij}. \tag{3.6}$$

The weighted or unweighted network representation (Fig. 3.1E,F) that is thus obtained from thresholding the similarity matrix $\mathbf{S}$ can now be characterized by means of local and global, unweighted and weighted network measures as they were introduced in Chap. 2.

In addition to setting a global threshold T based on the prescription of a desired link density d a climate network's adjacency matrix can also be obtained by defining individual per-link thresholds T_{ij} between nodes i and j. Even though rarely applied, this approach is particularly reasonable when associating the similarity values S_{ij} with a statistical significance that depends on the specific time series under study [Paluš et al., 2011]. In this case, individual thresholds T_{ij} are for example obtained from comparing the actual similarity estimation S_{ij} with that of time series generated from (computationally comparatively expensive) surrogate data techniques such as Fourier based surrogates [Schreiber and Schmitz, 2000, Theiler et al., 1992].

3.4. Generalization to coupled climate networks

Up to this point functional (climate) networks have been considered as a representation of the interdependencies within a single (climatic) observable that is given as a

set of N_x time series $X_i(t)$. However, Donges et al. [2011b] proposed a framework to extend this procedure to also study interactions between different observables, such as different layers of geopotential height fields located at different altitudes parallel to the Earth's surface. In this case, each layer or observable constitutes one of N_G subnetworks G_n with N_{x_n} nodes that form an interacting or coupled network as presented in Sec. 2.4. In the context of climate network analysis, these interacting networks have also been denoted as *coupled climate networks* [Donges et al., 2011b, Wiedermann et al., 2016a]. It must however be noted, that the usage of the term *coupled* in *coupled climate networks* does usually not imply the notion of any directionality or causal influence between the different fields or layers under study. It is simply meant to indicate the fact that the network is composed of more than a single climatic field.

In coupled climate networks, the sets of links E_{nn} indicate statistical similarities within each layer, i.e., subnetwork G_n, and are obtained in full analogy to the construction of a single-layer climate network as outlined in the previous Secs. 3.2 and 3.3. In addition, sets of links E_{nm} indicate statistical similarities between nodes in two subnetworks G_n and G_m, respectively.

In order to obtain a representation of a coupled climate network it is now required to obtain (i) the corresponding estimations of the $N_{x_n} \times N_{x_n}$ similarity matrices $\mathbf{S}_n$ between nodes within one layer or subnetwork G_n and (ii) the $N_{xn} \times N_{xm}$ similarity matrices $\mathbf{S}_{nm}$ indicating interdependencies between nodes i in subnetwork G_n and j in G_m. The entire $N_x \times N_x$ ($N_x = \sum_{n=1}^{N_G} N_{xn}$) *super*-similarity matrix $\mathbf{S}$ is then given as analogue to the adjacency matrix of an interacting network (Eq. 2.16) as

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_1 & \mathbf{S}_{1,2} & \cdots & \mathbf{S}_{1,N_G} \\ \mathbf{S}_{2,1} & \mathbf{S}_2 & \cdots & \mathbf{S}_{2,N_G} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_{N_G,1} & \mathbf{S}_{N_G,2} & \cdots & \mathbf{S}_{N_G} \end{pmatrix}. \tag{3.7}$$

Again, the actual choice of similarity estimators may depend on the specific application. However, as for the case of single-layer climate network most studies on coupled climate networks relied up to now on the Pearson correlation coefficient at zero lag ρ_{ij} (Eq. (3.3)) [Donges et al., 2011b, Feng et al., 2012, Wiedermann et al., 2016a]. In order to compute similarities, i.e, cross-correlation, between nodes i in subnetwork G_n and nodes j in G_m, the definition of ρ_{ij} in Eq. (3.3) is slightly refined such that,

$$\rho_{nm,ij} = \frac{\sum_t (X_i(t) - \overline{X}_i)(X_j(t) - \overline{X}_j)}{\sum_t (X_i(t) - \overline{X}_i)^2 \sum_t (X_j(t) - \overline{X}_j)^2}, \quad i = 1, \ldots, N_{xn},\ j = 1, \ldots, N_{xm}. \tag{3.8}$$

In order to again simultaneously account for positive and negative feedback cycles in

the climate system the corresponding entries $S_{nm,ij}$ of estimated similarity between nodes in the different layers are considered in their absolute fashion,

$$S_{nm,ij} = |\rho_{nm,ij}|. \tag{3.9}$$

From the super-similarity matrix $\mathbf{S}$ as given in Eq. (3.7) it is now again straightforward to obtain the coupled climate network's super-adjacency matrix by individually thresholding each of the *internal* and *cross*-similarity matrices $\mathbf{S}_n$ and $\mathbf{S}_{nm}$, respectively:

$$\mathbf{A} = \begin{pmatrix} \Theta(\mathbf{S}_1 - T_1) & \Theta(\mathbf{S}_{1,2} - T_{1,2}) & \cdots & \Theta(\mathbf{S}_{1,N_G} - T_{1,N_G}) \\ \Theta(\mathbf{S}_{2,1} - T_{2,1}) & \Theta(\mathbf{S}_2 - T_2) & \cdots & \Theta(\mathbf{S}_{2,N_G} - T_{2,N_G}) \\ \vdots & \vdots & \ddots & \vdots \\ \Theta(\mathbf{S}_{N_G,1} - T_{N_G,1}) & \Theta(\mathbf{S}_{N_G,2} - T_{N_G,2}) & \cdots & \Theta(\mathbf{S}_{N_G,N_G} - T_{N_G}) \end{pmatrix}. \tag{3.10}$$

If no node weights w_i are to be considered in the assessment of the coupled climate network, all entries on the main diagonal of $\mathbf{A}$, i.e, all possibly emerging self-loops, must additionally be removed and, hence, one sets $A_{ii} = 0 \; \forall \; i = 1, \ldots, N_x$. Otherwise, the derivation in Eq. (3.10) already denotes the extended adjacency matrix $\mathbf{A}^+$ that is needed for the assessment of n.s.i. network characteristics (Sec. 2.1).

The choice of individual thresholds T_n for each layer and T_{nm} between pairs of two layers accounts for possible heterogeneities in the distribution of similarity estimates, i.e., correlation coefficients, between the different layers. In particular, it might be that naturally the correlation between nodes in one layer is on average larger than between nodes in another layer, for example when comparing observables representing different climatic systems, such as ocean and atmosphere [Feng et al., 2012, Wiedermann et al., 2016a]. Furthermore, it is to be expected that correlations within one single layer are usually much stronger than correlations between different layers due to, e.g, the stratification of the atmosphere [Donges et al., 2011b, Wiedermann et al., 2016a]. Thus, it is advisable to compute *internal* thresholds T_n and *cross*-thresholds T_{nm} by fixing individual internal link-densities d_n (Eq. (2.17)) and cross-link densities d_{nm} (Eq. (2.18)), and computing the respective quantiles π_n and π_{nm} of each internal and cross-similarity matrix $\mathbf{S}_n$ and $\mathbf{S}_{nm}$, respectively.

This section concludes the necessary steps to construct a functional (climate) network from a data set of input time series. A visual summary of all individual processes that are involved in the construction of a climate network is given in Fig. 3.1. Starting from here, the following two chapters apply the presented framework to distinguish different phases of the El Niño Southern Oscillation based on the assessment of a set of single-layer functional climate networks that are computed at different time steps. The subsequent Chapter 6 then makes use of the extended framework to construct

and study coupled climate networks in order to discriminate different topological characteristics in northern hemispheric ocean-atmosphere interactions.

Chapter 4.

Discrimination of Eastern and Central Pacific ENSO phases

With the tools of network theory as presented in Chap. 2 and its specific refinements to construct network representations of the climate system (Chap. 3) at hand, this chapter applies this framework to highlight the ability of climate network analysis to consistently and objectively discriminate different types of spatially extended temperature signals in the evolution of the El Niño Southern Oscillation (ENSO). In particular, a sequence of climate networks is computed for different time steps and each networks' node splitting invariant (n.s.i.) link-weighted transitivity $\mathcal{T}^*$ is evaluated as a measure for the localized or dispersed spatial organization of ENSO's teleconnections. From there it is observed that $\mathcal{T}^*$ displays a sharp peak during certain ENSO periods, which can be attributed to so-called East Pacific (EP) ENSO events. In contrast, the absence of a significant signal in $\mathcal{T}^*$ indicates the presence of a Central Pacific (CP) ENSO event. In comparison with recent studies, this method confirms all EP and CP El Niños between 1951 and 2014 which have been commonly defined by Hendon et al. [2009], Hu et al. [2011], Kim et al. [2011], Kug et al. [2009], Larkin and Harrison [2005], Yeh et al. [2009] and Graf and Zanchettin [2012] and further provides a consistent assignment for those periods that were ambiguously classified.

Sec. 4.1 introduces the relevant climatic background together with a review of existing works regarding the discrimination of different ENSO phases as a basis to formulate the research question that is addressed in this chapter. The specific climate data as well as further details on the employed methodology are presented in Secs. 4.2 to 4.4. Sec. 4.5 presents the results of the climate network based discrimination together with a sensitivity analysis and compares the observations with recently proposed discriminations in the literature. The results of this study are then briefly summarized in Sec. 4.6.

For the sake of better readability, $\mathcal{T}^*$ is in this chapter being referred to as *transitivity* and the prefixes *n.s.i.* and *link-weighted* are omitted. Further the n.s.i. strength s_i^* is simply referred to as node strength, thus omitting the prefix *n.s.i.* for

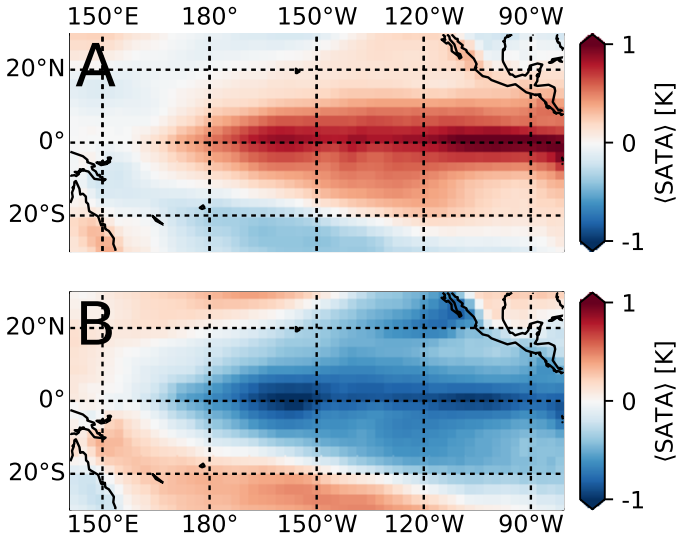

Figure 4.1.: Average daily surface air temperature anomalies (SATA) during (A) El Niño and (B) La Niña periods between 1948 and 2015.

brevity.

This chapter is based on the study that is presented in P6 [Wiedermann et al., 2016c].

4.1. Climatic background

The El Niño Southern Oscillation (ENSO), which is primarily observed over the tropical eastern Pacific Ocean, is one of the dominant modes of global climate variability [Trenberth, 1997]. Even though being spatially well confined it has been shown that ENSO has a large long-ranging impact on regional climate variability along different parts of the Earth [Ashok et al., 2007, Camberlin et al., 2001, Chiew et al., 1998]. This effect is often referred to as teleconnectivity. ENSO itself is characterized by an irregularly periodic variation in surface winds and sea surface temperatures (SST) with an average periodicity of seven to eight years [Trenberth, 1997]. It displays three major stages denoted as El Niño, La Niña and normal conditions. All three phases are related with variations in the Walker circulation that itself is caused by a pressure gradient along the Equator [Bjerknes, 1969].

During normal ENSO conditions the Walker circulation is characterized by westerly winds along the Equatorial Pacific that originate from a high pressure system over the East Pacific as well as a low pressure system over Indonesia. In particular, the northward flowing Humboldt current transports relatively cold sub-surface water from the Southern (Antarctic) Ocean northwards to the South American west

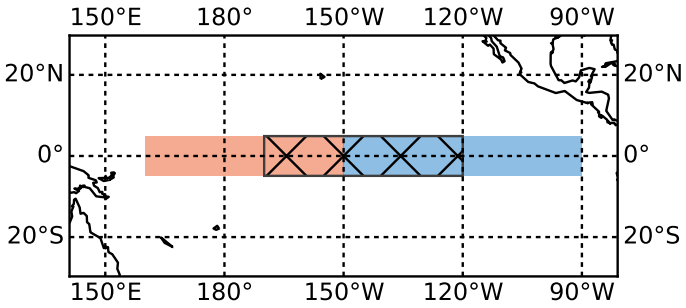

Figure 4.2.: Spatial extent of the Nino3 region (5°N to 5°S and 150°W to 90°W, blue shaded area), the Nino4 region (5°N to 5°S and 160°E to 150°W, red shaded area) and the Nino3.4 region (5°N to 5°S and 170°W-120°W, hatched area) in the Equatorial Pacific.

coast where it is transported to the surface via up-welling. This process then cools the surface of the eastern Pacific Ocean. The trade winds induced by the Walker circulation then transport this cold water westwards where it slowly warms in the process and ultimately causes the western Pacific Ocean to be significantly warmer than the East Pacific.

In case the Walker circulation weakens or reverses, less cold water is transported westwards and the upwelling of cold subsurface water occurs less. An input of less cold water causes an anomalous warming of sea surface temperature in the East Pacific [Allan et al., 1997]. These positive temperature anomalies indicate El Niño conditions and their corresponding highly correlated spatial extent of surface air temperature anomalies (SATA) is displayed in Fig. 4.1A.

In contrast, a stronger than average Walker circulation enhances the upwelling of cold subsurface water along the South American west coast causing negative sea surface temperature anomalies in this region. Such cold conditions are denoted as La Niña periods [Allan et al., 1997]. Fig. 4.1B again illustrates the spatial extent of these conditions in terms of surface air temperatures anomalies.

Several indicators for the presence of either of these three possible states of ENSO exist. One, that is based on atmospheric observations, is the Southern Oscillation Index (SOI) that depends on the difference in sea level pressure between Tahiti and Darwin (Australia) [e.g., Trenberth, 1984]. Positive values of the SOI indicate La Niña and negative values indicate El Niño conditions.

A variety of further commonly used indices is based on oceanic observations. Specifically, the Nino3 index is computed as average sea surface temperature anomalies (SSTA) in the region between 5°N to 5°S and 150°W to 90°W (blue shaded are in Fig. 4.2). This region was once the primary focus for El Niño observations [Trenberth, 1997]. Starting from there it has been observed that the key region for ENSO-related ocean-atmosphere interactions is in fact located further west in the Pacific Ocean.

Therefore, the Nino4 region was defined as the area from 5°N to 5°S and 160°E to 150°W which captures SST anomalies in the central Equatorial Pacific (red shaded area in Fig. 4.2) [Trenberth, 1997]. However, this region was found to display only a low degree of variance in the observations of SST anomalies. Ultimately, the Nino3.4 region was defined as the area covering 5°N to 5°S and 170°W-120°W and is utilized since as one primary focus for monitoring the evolution of ENSO (hatched area in Fig. 4.2) [Trenberth, 1997]. One way to define El Niño and La Niña periods from SST observations in the Nino3.4 region is to evaluate the temporal evolution of the Oceanic Niño Index (ONI) that is provided by the Climate Prediction Center of the National Oceanic and Atmospheric Administration. By the time of conducting the study, this index covered the time between 1950 and 2015 and it is computed as the 3-month running mean SSTA in the Nino3.4 region with respect to centered 30-year base periods that are updated every 5 years. The ONI identifies El Niño (La Niña) episodes if its values exceed (fall below) a threshold of 0.5K (-0.5K) for at least 5 consecutive months, yielding 22 (18) El Niño (La Niña) episodes between 1951 and 2014 (Fig. 4.4A).

Both, El Niño and La Niña, have been shown to have a long-ranging effect on climate variability at different regions along the Earth. Examples include an increase in precipitation over Northern China during El Niño periods [Zhang et al., 1999] contrasted by anomalously dry conditions over, e.g., Australia [Chiew et al., 1998]. In contrast, the presence of La Niña periods have for example been linked with an increase in precipitation over Northern Europe and corresponding negative rainfall anomalies in Southern Europe [Pozo-Vázquez et al., 2005]. Other studies report an increase in precipitation over the South African subcontinent that is related with the presence of La Niña conditions [Nicholson and Selato, 2000]. Further, El Niño and La Niña have been linked to dynamics related with the Indian Summer Monsoon, especially with a focus on the strengthening or weakening of its related precipitation signals [Kumar et al., 2006, Zhang et al., 1996]. Accounting for all such global influences, these and other observed long-ranging impacts of ENSO are commonly referred to as its associated *teleconnections* [Glantz et al., 1991].

In addition to a discrimination of ENSO into El Niño and La Niña periods, it has been reported that especially the El Niño phase further exhibits two distinct types characterized by different spatial patterns of SSTA or SATA [e.g., Ashok et al., 2007, Kao and Yu, 2009, Kug et al., 2009, Yeh et al., 2009]. The first type, the classic or Eastern Pacific (EP) El Niño [Harrison and Larkin, 1998, Rasmusson and Carpenter, 1982], is characterized by strong positive SSTA or SATA close to the western coast of South America (such as during the El Niño in 1997/1998, Fig. 4.3A). The second type, referred to as El Niño Modoki or Central Pacific (CP) El Niño by different authors, exhibits the strongest anomalies close to the dateline (such as during the El Niño in 1994/1995, Fig. 4.3B). Both types cause different impacts on the global

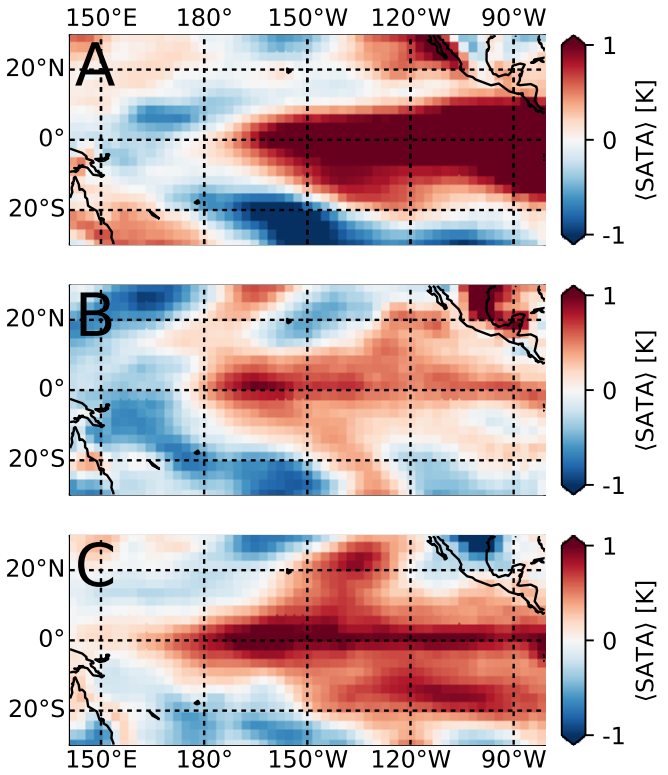

Figure 4.3.: Average daily surface air temperature anomalies for the three different El Niños during (A) 1997/1998, (B) 1994/1995 and (C) 1986/1987.

climate system, such as increased rainfall over northern and eastern Australia during CP El Niños [Ashok et al., 2007, Taschetto and England, 2009] contrasted by a rainfall reduction over eastern Australia during EP El Niños [Chiew et al., 1998]. Thus, a proper discrimination of these types provides key information to assess El Niño's possible impacts on other climate subsystems. While recent literature shows a large agreement on the classification of many El Niños, contradictory classifications arise in certain years such as, e.g., 1986/1987, which has been classified as mixed [Kug et al., 2009], EP [Hu et al., 2011, Kim et al., 2011, Yeh et al., 2009] or CP [Graf and Zanchettin, 2012, Hendon et al., 2009, Larkin and Harrison, 2005]. For this particular period strong SSTA or SATA are observed along the entire equator between the dateline as typical for the CP type and the coast of South America as typical for the EP type (Fig. 4.3C). Hence, it is immediately apparent that a discrimination based on (average) spatial anomalies alone may not be sufficient to provide a proper discrimination of the different types.

In fact, when reviewing existing studies [Graf and Zanchettin, 2012, Hendon et al., 2009, Hu et al., 2011, Kim et al., 2009, 2011, Kug et al., 2009, Larkin and Harrison, 2005, Yeh et al., 2009], 8 out of 19 El Niño events between 1957 and 2010 have not been classified in agreement. These mismatches possibly arise since most dis-

crimination schemes utilize the same observable (mostly SST), but apply different derived characteristics. One such characteristic is for example the ENSO Modoki Index (EMI) [Ashok et al., 2007], which is defined as a weighted sum over three area-averaged SSTA values in the Pacific Ocean. Further discrimination schemes are based on a combined assessment of the Nino3 and Nino4 index [Hu et al., 2011, Kim et al., 2011] or empirical orthogonal function (EOF) analysis [Graf and Zanchettin, 2012, Kao and Yu, 2009]. Especially the latter requires some manual thresholding of the EOF's time evolution which may result in ambiguous classifications that strongly depend on the choice of the threshold.

To provide a consistent and systematic discrimination, this chapter proposes a method to distinguish the two different El Niño types based on the assessment of time evolving complex climate networks [Radebach et al., 2013]. For this purpose the concepts of functional climate network analysis as presented in Chap. 3 are extended such that a sequence of climate networks is computed for different time steps which enables to track the temporal evolution of topological network characteristics as the system evolves in time.

In the past, climate networks have already been applied intensively to study ENSO's corresponding teleconnections [Gozolchiani et al., 2011, Tsonis and Swanson, 2008, Tsonis et al., 2008], its effect on other climatic subsystems [Gozolchiani et al., 2008, Martin et al., 2013] and the dynamics of its related oceanic wave dynamics [Wang et al., 2016]. Additionally, climate network approaches allowed for successfully forecasting El Niño by assessing the strength of linkages in the equatorial Pacific [Ludescher et al., 2013, 2014].

Radebach et al. [2013] systematically studied the temporal evolution of a global climate network in a spatially explicit way and linked the resulting variability of its topology to the presence of the two different El Niño types. Following upon these results, this chapter presents a classification scheme that allows for an objective discrimination between EP and CP El Niños. While most previous studies on El Niño classification focus on climate variability only within the equatorial Pacific, climate networks particularly acknowledge the global impact of ENSO by accounting for the correlation structure of global SATA. This variable is highly affected by El Niño [Yamasaki et al., 2008] and is, in contrast to SST, available homogeneously sampled for the entire globe.

4.2. Data

The surface air temperature anomalies (SATA) are derived from daily global surface air temperature (SAT) data provided by the NCEP/NCAR reanalysis [Kistler et al., 2001] with a spatial resolution of 2.5° in longitudinal and latitudinal direction covering the same time period as the ONI. All 288 grid points located at the poles and

all leap days are removed. The data is anomalized in accordance with the definition of the ONI by subtracting from the time series at every grid point the corresponding long-term annual cycle. These annual cycles are computed over the same 30-year base periods that are updated every 5 years as used in the definition of the ONI[1]. Due to the lack of data before 1948 and after 2015, the years 1951 to 1965 are anomalized by the same base period (1951–1980) as the years 1965 to 1969. Similarly, the years 2005 to 2015 are anomalized by the 1986 to 2015 base period. This anomalization process ensures that once defined anomalies and ENSO periods are not altered by the addition of more recent data.

Finally, $N_x = 10,224$ daily SATA time series $X_i(t)$ with $N_t = 23,360$ temporal sampling points each are obtained and further utilized for the construction of evolving climate networks.

4.3. Network construction

Following the framework of evolving climate networks [Radebach et al., 2013] a sequence of networks G_n is constructed from running-window cross-correlation matrices between all pairs of SATA time series. A window n is characterized by its size ω and offset o to the previous window which is chosen here as $o = 30$ days and $\omega = 365$ days. This particular choice ensures that each window covers at least the entire duration of an El Niño or La Niña episode. For each window n the truncated time series $\{X_{n,i}(t_n)\}$, $t_n = no, no+1, \ldots, no+\omega-1$ is obtained and from there the resulting $N_x \times N_x$ cross-correlation matrix $\mathbf{C}_n$ is computed. In accordance with previous studies that utilized either monthly [Donges et al., 2009b, Paluš et al., 2011, Tsonis et al., 2006] or daily [Radebach et al., 2013] data, the linear Pearson correlation ρ_{ij} at zero lag as given in Eq. (3.3) is used to compute the entries of each correlation matrix $\mathbf{C}_n$. To account, for positive and negative feedback cycles in the climate system (Sec. 3.2), the similarity matrices $\mathbf{S}_n$ that are utilized for the network construction are computed according to Eq. (3.4) as the absolute value of pairwise cross-correlation ρ_{ij}. Thus, the entries $S_{n,ij}$ of $\mathbf{S}_n$ read

$$S_{n,ij} = |\rho_{n,ij}|. \tag{4.1}$$

To construct each of the evolving climate networks G_n from the corresponding similarity matrices $\mathbf{S}_n$, a prescribed link density of $d = 0.005$ is fixed and individual thresholds T_n for each matrix are obtained. This particular choice of d roughly corresponds to the fraction of nodes that are situated inside the Nino3.4 region (Fig. 4.2) and yields thresholds T_n in the range of 0.53 to 0.65 which are significant above the 99% significance level according to a standard student's t-test.

[1] `www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_change.shtml`

Taken together, the construction of evolving climate networks depends on only two parameters: Window size ω and link density d. Compared to, e.g., the EMI [Ashok et al., 2007] which requires at least the weights of its three contributions to be fixed, the number of parameters in the network-based framework is (i) of comparable order and (ii) each parameter can be selected in a meaningful way according to ENSO's temporal (window size ω) and spatial extent (link density d).

Binarizing the similarity matrix $\mathbf{S}_n$ to a link-unweighted adjacency matrix $\mathbf{A}_n$ would neglect valuable information on the varying strength of correlation between connected grid points. Therefore, the link-weighted adjacency matrices $\mathbf{W}_n$ with corresponding entries $W_{n,ij}$ are computed according to Eq. (3.6) as

$$W_{n,ij} = \Theta(|\rho_{n,ij}| - T_n)(1 - \delta_{ij})|\rho_{n,ij}|. \tag{4.2}$$

4.4. Quantification of the network topology

As discussed in Sec. 4.1, ENSO has a global impact on the climate system manifested by long-ranging teleconnections with different regions of the Earth [Held et al., 1989, Neelin et al., 2003, Trenberth, 1997] which, in the context of climate networks can be regarded as mediators of variations and fluctuations [Runge et al., 2015, Tsonis et al., 2008]. Thus, ENSO and its teleconnections cause a spatial organization of high co-variability along the Earth's surface which is reflected in the resulting climate network. The degree of this organization can be well quantified by the transitivity $\mathcal{T}^*$ as given in Eq. (2.14) [Tsonis et al., 2006]. It gives the link- and node-weighted fraction of completely linked triples of nodes and measures how strongly the correlation in the system under study or subsets thereof is spatially organized (high values) or dispersed (low values). $\mathcal{T}^*$ thus serves as a good discriminator between phases of strong localization and high dispersion in the global teleconnectivity of evolving climate networks [Radebach et al., 2013]. As EP and CP El Niños have been shown to display different characteristics in their associated teleconnections [Ashok et al., 2007], $\mathcal{T}^*$ is expected to respond differently to the presence of either of the two types.

4.5. Results

El Niño and La Niña periods are defined by specific values of the ONI that was at the time of conducting this study available for the period from 1950 to 2015 (Fig. 4.4A). As the initial and final year of this data set include only incomplete information on the 1951 La Niña and the 2015 El Niño, this work is restricted to the period from 1951 to 2014 (Fig. 4.4A). Subsequently, a total number of $n = 733$ evolving climate networks is constructed and their transitivity $\mathcal{T}_n^*$ and node strength $s_{n,i}$ are computed. The end point of each window marks the time at which the two measures

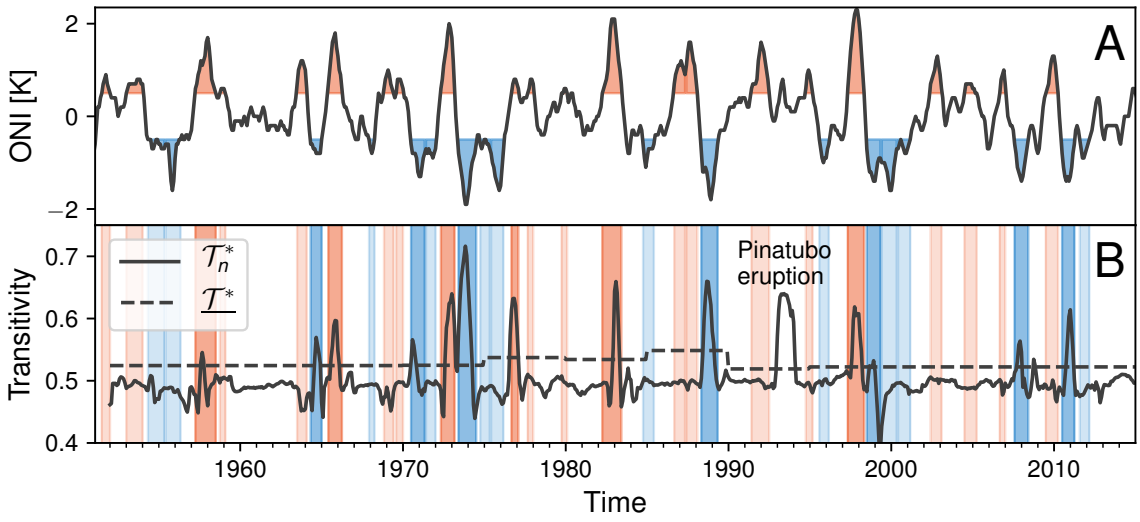

Figure 4.4.: (A) The ONI with El Niño (La Niña) periods marked in red (blue). (B) Time evolution of the evolving climate network's transitivity $\mathcal{T}_n^*$. The dashed horizontal line indicates the baseline transitivity $\underline{\mathcal{T}}^*$. Colored areas highlight El Niño and La Niña periods. Darker coloring indicates those periods where $\mathcal{T}_n^*$ exceeds $\underline{\mathcal{T}}^*$ and that are thus classified as EP type. (from Wiedermann et al. [2016c])

are evaluated.

Transitivity

Figure 4.4B shows the evolution of the transitivity $\mathcal{T}_n^*$. Except for one case with several 12-month time windows ending in 1993, which reflects large-scale spatially coherent cooling after the Mount Pinatubo eruption in 1991 [McCormick et al., 1995, Radebach et al., 2013], peaks in $\mathcal{T}_n^*$ coincide exclusively with distinct ENSO episodes. As shown by Radebach et al. [2013], the presence of an EP El Niño likely coincides with strong signals in (for their case unweighted) transitivity, while no distinct signal is present during CP El Niños. However, no quantitative criterion for this discrimination has been given so far.

To give an objective definition of a *strong* transitivity, a threshold value $\underline{\mathcal{T}}^*$ is defined above which $\mathcal{T}_n^*$ is considered to display a peak. In particular, $\underline{\mathcal{T}}^*$ is obtained adaptively as the transitivities of climate networks constructed for the same 30-year periods that were used for the anomalization of the SAT data and the derivation of the ONI. Thus, all values of $\mathcal{T}_n^*$, e.g., during the period 1975–1979, are compared with a baseline transitivity $\underline{\mathcal{T}}^*$ computed for a climate network covering the 30-year period of 1961–1990 (dashed line in Fig 4.4B). This procedure follows the definition of the ONI and one may thus directly interpret $\underline{\mathcal{T}}^*$ as a representation of the long-term average spatial organization in the global climate network. Adaptively updating $\underline{\mathcal{T}}^*$ every 5 years automatically accounts for possible effects of long-term trends

	Kug et al. (2009)	Kim et al. (2011)	Hu et al. (2011)	Larkin et al. (2005)	Hendon et al. (2009)	Graf et al. (2012)	Yeh et al. (2009)	Kim et al. (2009)	Common	Full results
1953/1954	-	-	-	-	-	-	-	-	-	CP
1957/1958	-	-	**EP**	**EP**	-	**EP**	**EP**	**EP**	**EP**	EP
1958/1959	-	-	-	-	-	-	-	-	-	CP
1963/1964	-	-	-	**CP**	-	**CP**	EP	EP	-	CP
1965/1966	-	-	**EP**	**EP**	-	**EP**	**EP**	**EP**	**EP**	EP
1968/1969	-	-	**CP**	**CP**	-	**CP**	**CP**	-	**CP**	CP
1969/1970	-	-	EP	EP	-	-	EP	**CP**	-	CP
1972/1973	**EP**	**EP**	**EP**	**EP**	-	**EP**	**EP**	**EP**	**EP**	EP
1976/1977	**EP**	**EP**	-	**EP**	-	**EP**	**EP**	**EP**	**EP**	EP
1977/1978	**CP**	**CP**	-	**CP**	-	**CP**	**CP**	-	**CP**	CP
1979/1980	-	-	-	-	-	-	*	-	-	CP
1982/1983	**EP**	**EP**	**EP**	**EP**	**EP**	**EP**	**EP**	**EP**	**EP**	EP
1986/1987	*	EP	EP	**CP**	**CP**	**CP**	EP	-	-	CP
1987/1988	*	-	**CP**	EP	EP	-	EP	EP	-	CP
1991/1992	*	EP	EP	EP	**CP**	**CP**	EP	**CP**	-	CP
1994/1995	**CP**	**CP**	**CP**	**CP**	**CP**	**CP**	**CP**	**CP**	**CP**	CP
1997/1998	**EP**	**EP**	**EP**	**EP**	**EP**	**EP**	**EP**	**EP**	**EP**	EP
2002/2003	**CP**	**CP**	**CP**	EP	**CP**	**CP**	*	**CP**	-	CP
2004/2005	**CP**	**CP**	-	-	**CP**	**CP**	**CP**	**CP**	**CP**	CP
2006/2007	-	EP	**CP**	**CP**	-	-	EP	-	-	CP
2009/2010	-	**CP**	-	-	-	**CP**	-	-	**CP**	CP
TPR	1.0	0.57	0.62	0.6	0.67	1.0	0.5	0.75	1.0	

Table 4.1.: Recent classifications of El Niño phases into CP and EP episodes. Asterisks denote mixed or undefined states. A hyphen denotes that no classification was performed for the specific year. Bold letters denote events where the network-based classification is in agreement with the reference. The last row summarizes the true positive rate (TPR) of the network-based formalism. The second-last column indicates the largest overlap between all references and the last column summarizes the classification obtained from the network-based approach.

imprinting on the network statistics and the definition of $\underline{\mathcal{T}}^*$ for periods in the past is not affected by the addition of more recent data.

Six El Niño periods are detected during which $\mathcal{T}_n^*$ exceeds $\underline{\mathcal{T}}^*$ (dark red areas in Fig. 4.4B) corresponding to the El Niños of 1957, 1965, 1972, 1976, 1982 and 1997. For all other El Niños $\mathcal{T}_n^*$ stays below $\underline{\mathcal{T}}^*$. In the scope of the framework presented here, it is thus proposed to classify the first case as EP and the second case as CP events (respective dark and light red areas in Fig. 4.4B).

For comparison, the proposed classifications of El Niño phases into EP and CP types from eight recent studies [Graf and Zanchettin, 2012, Hendon et al., 2009, Hu et al., 2011, Kim et al., 2009, 2011, Kug et al., 2009, Larkin and Harrison, 2005, Yeh et al., 2009] are summarized in Tab. 4.1. To quantify the consistency of the network-based discrimination, a true positive rate (TPR) is defined as the fraction of EP El Niños in each study that are also detected by the network based discrimination scheme. Accordingly, the false positive rate (FPR) is the fraction of CP El Niños in each study that the transitivity $\mathcal{T}_n^*$ classifies as EP type. With respect to all references a FPR of zero is obtained meaning that no CP event is falsely classified as an EP event by the network-based approach. The TPR for each reference is presented in the last row of Tab. 4.1. Its values vary between 1 for the comparison with results from Graf and Zanchettin [2012] and Hu et al. [2011] and 0.5 for the

comparison with discriminations provided by Yeh et al. [2009].

Furthermore, it has to be noted that among all references 8 out of 19 events are not classified in agreement. Taking only the mutual agreement between all references as a basis for testing, all past classifications are confirmed by the network-based discrimination (second-last column in Tab. 4.1). To provide results for the eight ambiguously defined periods, the network-based classification for all El Niños is given in the last column of Tab. 4.1.

Notably, the network-based framework achieves the largest consistency with the results from Graf and Zanchettin [2012] which are obtained from an empirical orthogonal function (EOF) analysis, a framework that, like the presented method, is based on the evaluation of cross-correlations between different grid points. This methodological congruence may explain the good agreement between the results and further confirms the validity of the network-based approach. However, by utilizing the transitivity $\mathcal{T}_n^*$ instead of EOFs the entire spatial structure of the underlying covariance patterns is reduced to a single index. Its evaluation does not rely on any visual inspection, but provides an objective binary classification depending on whether or not the short-term transitivity $\mathcal{T}_n^*$ exceeds its long-term baseline $\underline{\mathcal{T}}^*$.

Repeating the analysis for La Niña periods uncovers 7 EP (1964, 1970, 1973, 1988, 1998, 2007, 2010) and 11 CP (1954, 1955, 1967, 1971, 1974, 1975, 1984, 1995, 2000, 2001 and 2011) periods (dark (EP) and light (CP) blue areas in Fig. 4.4B). Even though references providing actual discriminations of the different La Niña years are scarce, two recent works were compiled that reported the EP La Niñas of 1964 and 1970 [Yuan and Yan, 2012] and CP La Niñas of 1975, 1984, 2000, 2001, and 2011 [Tedeschi et al., 2013, Yuan and Yan, 2012]. Thus, future work should further evaluate the discrimination of La Niña periods proposed by the network-based method.

Node strength

To further consolidate the findings presented in the previous section the average node strengths during normal periods as well as the four different ENSO stages, as classified by the network-based approach (Tab. 4.1), are computed. Generally, the spatial patterns of node strength s_i^* (Eq. (2.6)) are expected to display a large degree of similarity with the first empirical orthogonal function [Donges et al., 2015b], making it generally possible to relate results obtained from climate network analysis with recent works that employ tools from classical statistical climatology.

First, the average node strength during normal periods $s_{B,i}^*$ is derived from the six networks that are used to define $\underline{\mathcal{T}}^*$ (Fig. 4.5A). The highest values of $s_{B,i}^*$ are obtained in the equatorial Pacific which again highlights ENSO's importance in the global climate network. Additionally, the average node strength $s_{normal,i}^*$ is computed over all *normal* periods, i.e., those periods where neither El Niño or La

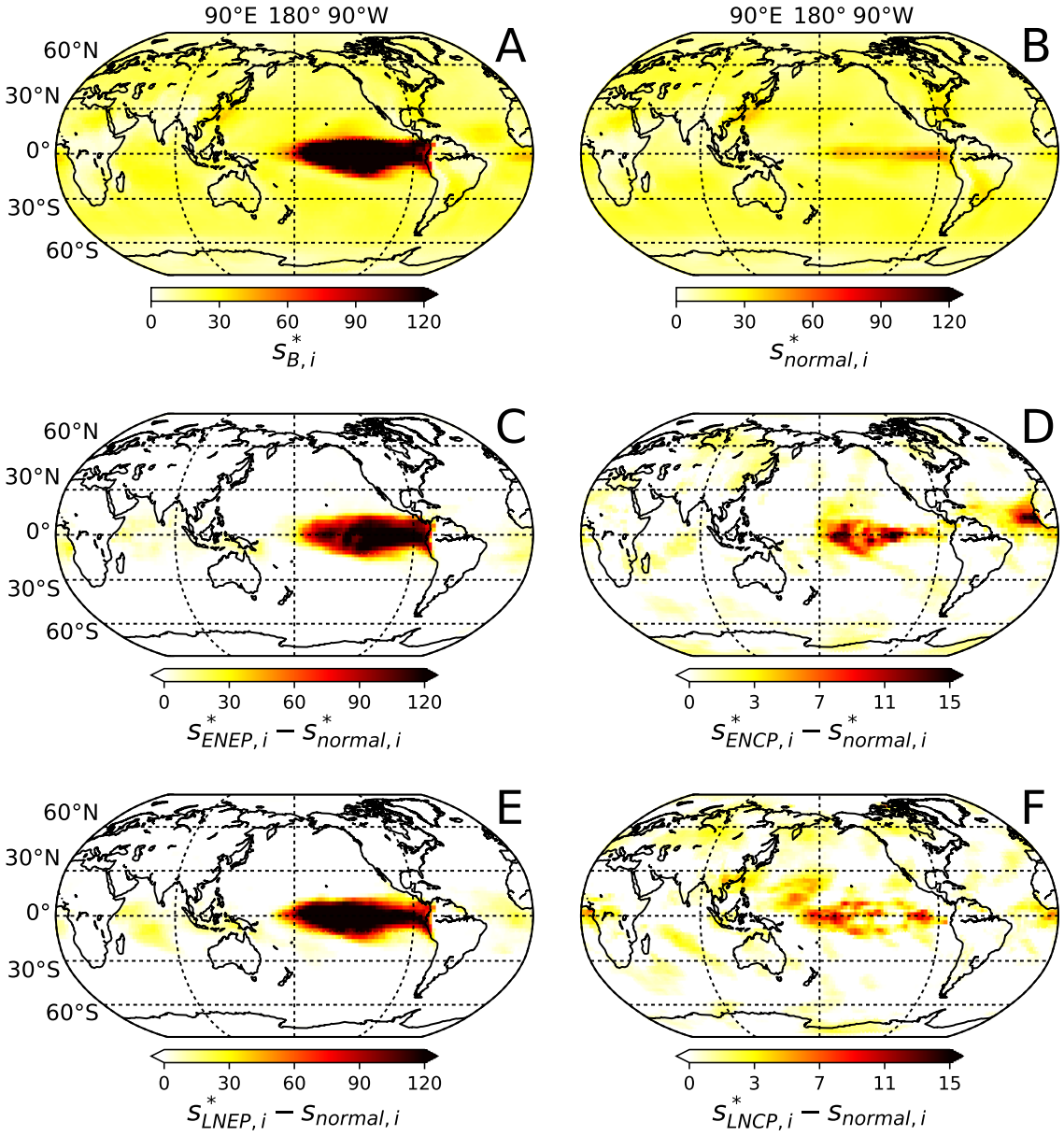

Figure 4.5.: (A) Average strength of nodes in the baseline climate networks. (B) Average node strength of the evolving networks during normal periods. (C) Differences between the average node strength during El Niño periods that are classified as EP type and the average node strength during normal periods. (D) The same as in (C) for El Niño periods that are classified as CP type. (E,F) The same as in (C,D) for La Niña periods. Note the different color ranges in (C,E) and (D,F), respectively (adapted from Wiedermann et al. [2016c]) .

Niña are present (Fig. 4.5B). As by its definition the effect of ENSO is reduced and $s^*_{normal,i}$ displays comparably low values as well as a relatively homogeneous distri-

bution across the entire globe when compared to $s^*_{B,i}$. Ultimately, the average node strength $s^*_{ENEP,i}$ ($s^*_{ENCP,i}$) is computed over all El Niño periods that the network-based method classifies as EP (CP) type (see also Fig. 4.4B and Tab. 4.1). To investigate the deviation from the normal state during either of the two periods their respective differences from $s^*_{normal,i}$ are displayed in Fig. 4.5C,D. For EP El Niños (Fig. 4.5C) an expected maximum is observed in the equatorial East Pacific, which is the typical ENSO-related pattern known from classical EOF analysis [Johnson, 2013]. For CP El Niños this pattern weakens and the maximum shifts westward towards the dateline. Such a pattern has again been observed in the corresponding EOFs as well [Johnson, 2013]. However, $s^*_{ENCP,i}$ only differs from $s^*_{normal,i}$ to a small amount (Fig. 4.5D) which again suggests that during CP El Niños the evolving climate networks exhibit a similar state as during normal periods. Similar average quantities, $s^*_{LNEP,i}$ and $s^*_{LNEP,i}$, are computed for La Niña events and their deviations from the normal state are again displayed in Fig. 4.5E,F. The observed patterns are quantitatively and qualitatively similar to those obtained for El Niño, which highlights the symmetry of the two ENSO phases. Even though a similarly thorough comparison with existing literature is not yet possible for La Niña, the high congruence between $s^*_{ENEP,i}$ and $s^*_{LNEP,i}$ ($s^*_{ENCP,i}$ and $s^*_{LNCP,i}$) suggests that the network-based discrimination scheme provides reasonable results for La Niña phases as well.

Robustness

To evaluate the robustness of the obtained results with respect to the window size ω and link density d, both parameters are varied individually and the difference between the TPR and FPR is assessed when testing the network-based classification against the largest overlap in the literature (second-last column in Tab. 4.1). This score takes its maximum value of 1 if the transitivity $\mathcal{T}^*_n$ confirms the literature's classification of each event and is lower otherwise. Figure 4.6A (Fig. 4.6B) shows the score for different ω (or d) with fixed $d = 0.005$ ($\omega = 365$ days). The highest scores are obtained for window sizes between $\omega = 365$ and $\omega = 547$ days and link densities between $d = 0.005$ and $d = 0.0075$. Shorter window sizes cause a reduction of the score as the windows become too small to sufficiently cover the temporal extent of an ENSO episode. For larger window sizes the effect of ENSO is suppressed by including too many of the normal periods into each window. The link density of $d = 0.005$ was initially chosen as it roughly corresponds to the fraction of nodes located inside the Nino3.4 region (Fig. 4.2). Smaller values cause the network to be only composed of highly correlated trivial nearest-neighbor connections and teleconnections with comparably lower pairwise cross-correlation vales are not captured. In contrast, larger values result in too many trivial links alongside those attributed to the effects of ENSO.

Generally, the score varies smoothly along the range of parameters and shows

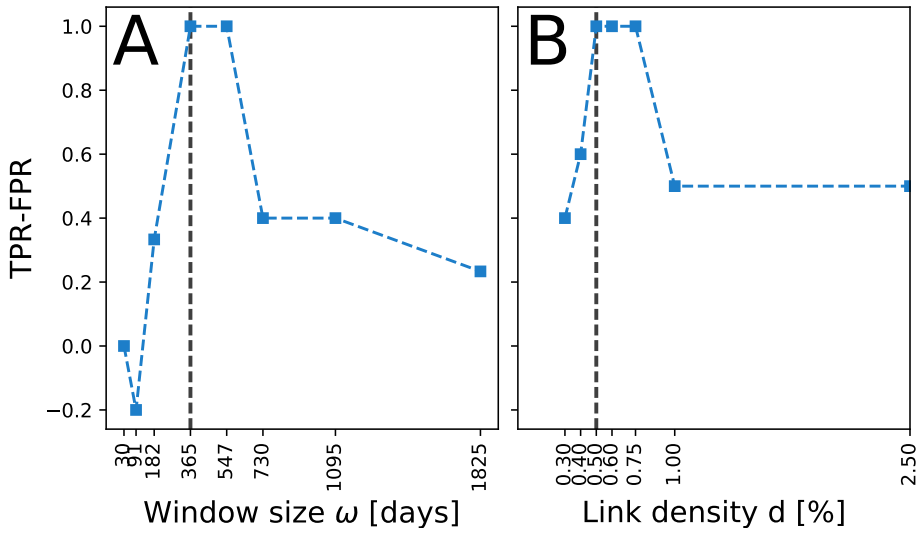

Figure 4.6.: (A) Difference between true positive rate (TPR) and false positive rate (FPR) for classifications obtained from the network-based transitivity index and the largest overlap between all references in Tab. 4.1 for different window sizes ω and fixed link density $d = 0.005$. (B) The same as in (A) for different link densities d and fixed window size $\omega = 365$ days. Dashed lines indicate the choices of parameters that yield the results in Fig. 4.4B and Fig. 4.5 (adapted from Wiedermann et al. [2016c]) .

maximum values for their initial choices. Thus, the results presented in the previous sections are considered to be sufficiently robust.

4.6. Summary

This chapter introduced an index based on evolving climate networks to objectively discriminate between EP and CP types of El Niño and La Niña episodes. It relies on the evolution of the networks' transitivity, measuring spatial localization and dispersion of strong cross-correlations between different grid points in a global SATA field. If this index peaks during a distinct ENSO phase, it detects the presence of an EP type event. In contrast, the absence of a remarkable signal during an ENSO period indicates CP type events. From the climate network perspective this indicates an increased localization and clustering of teleconnections during EP phases in comparison with CP and normal phases where teleconnections seem to appear more dispersed. Specifically, the proposed method does not require any visual inspection or manual thresholding of observed patterns but objectively categorizes ENSO phases into different types by inter-comparing the networks' short-term ($\mathcal{T}_n^*$) and long-term states ($\underline{\mathcal{T}}^*$).

In comparison with eight recent works on El Niño classification the proposed

network-based method confirms the classification of years that all references have in common and provides a discrimination for those years that were so far ambiguously defined. Unlike approaches based on the evaluation of (average) SST fields or first-order statistics thereof the transitivity produces a sharp and distinct signal, and thus provides a clear distinction between the two types of El Niño episodes.

Even though references are scarce, the findings also confirm different recently reported EP and CP La Niña periods and show that the network-based discrimination scheme is applicable to this negative phase of ENSO as well.

Future work should investigate more thoroughly the spatial distribution of links in the evolving climate networks during different ENSO stages to gain a more systematic understanding of the physical mechanisms behind the observed differences in transitivity. Moreover, being automated and objective, the proposed framework allows for a systematic evaluation of climate model simulations and could be used to investigate potential changes in the projected frequency of the two ENSO flavors in the future, e.g., due to anthropogenic global warming [Yeh et al., 2009].

In summary, a discrimination of El Niño and La Niña episodes based on the assessment of functional climate networks is a meaningful complement to existing frameworks as it (i) proposes objective classifications where former work yielded ambiguous results and (ii) depends only on a low number of parameters which can be estimated from ENSO's well-studied spatio-temporal extent, while (iii) showing no larger computational complexity than EOF-based methods. Further, it was shown that the exact choice of parameters does not affect the outcome of the analysis as long as they are varied over a climatically reasonable range.

With the network-based discrimination scheme at hand it is now straightforward to assess possible impacts of each of the four different ENSO types (East Pacific and Central Pacific El Niño and La Niña) on different climatic observables. Among many others, such as responses in temperature, one crucial process here is ENSO's impact on extreme precipitation signals [Dai and Wigley, 2000, Wiedermann et al., 2017b]. The next chapter presents such a comprehensive analysis by utilizing the network based transitivity index as a binary indicator for the presence of either ENSO phase and evaluates its imprints on spatially coherent patterns of global extreme precipitation.

Chapter 5.

Extreme climate responses to different ENSO phases

The previous chapter introduced an index to discriminate different phases of the El Niño Southern Oscillation (ENSO) based on assessing the corresponding functional evolving climate network's transitivity. This chapter now utilizes those findings to investigate extreme climatic responses in terms of precipitation signals to the thus discriminated four different types of ENSO. The impacts of specifically El Niño's two flavors, East Pacific (EP) and Central Pacific (CP) El Niño, have been studied intensively in recent years, mostly by applying linear statistical or composite analyses. The used techniques, however, focus on average spatio-temporal patterns of climate variability and do not allow for a specific assessment of related extreme impacts. Hence, the present chapter specifically focuses on the latter by using *event coincidence analysis* to study the differential imprints of EP and CP types of both, El Niño and La Niña on global extreme precipitation patterns which severely affect, e.g., agricultural and biomass production or human health. The work reveals that EP events usually coincide with spatially coherent patterns of seasonal hydro-meteorological extremes, while more spatially dispersed patterns emerge for CP phases, respectively. This aligns well with the assumption from the previous chapter on the localization and dispersion of ENSO's teleconnections during EP and CP phases, respectively, which form the basis of the network-based discrimination scheme. The analysis presented here recaptures previously reported interrelations and uncovers further regional extremes arising along with different ENSO phases that have not been reported so far. The results imply that a proper discrimination of El Niño and La Niña into its distinct phases are crucial for anticipating meteorological as well as related socio-ecological impacts.

Sec. 5.1 reviews the current state of research on ENSO related extreme climate responses and introduces the specific research task that is addressed with this work. Sections 5.2 and 5.3 discuss the data sets that are used in addition to the ones employed in the previous chapter as well as all necessary steps of preprocessing. From this, Sec. 5.4 introduces the mathematical concepts of *event coincidence analysis*

which is applied to statistically determine interrelations between ENSO and global extreme precipitation signals. The results of the study are presented in Sec. 5.5. It opens by evaluating the effects of the canonical (EP) El Niño on extreme precipitation recapturing a variety of previously reported interrelations. This demonstrates the validity of the presented approach. The focus is then shifted towards CP El Niños and differences with their EP counterparts are discussed. Ultimately, the study investigates also La Niña periods and shows that most extreme rainfall responses are attributed to the EP type of ENSO's negative phase. In the light of recent discussions on the existence of two types of La Niña [Kug and Ham, 2011] it seems therefore meaningful to distinguish one type that significantly affects global precipitation signals and one that shows hardly any spatially coherent impact. Wherever appropriate, possible socio-ecological consequences of the observed seasonal precipitation extremes are discussed as well. Finally, Sec. 5.6 provides a summary and the conclusions of this chapter and provides suggestions for future research.

This chapter is primarily based on the study that is presented in P12 [Wiedermann et al., 2017b].

5.1. Climatic background

El Niño and La Niña are both known to trigger climatic responses in various parts of the Earth, an effect which was introduced in Chap. 4 as teleconnectivity [Neelin et al., 2003, Trenberth, 1997]. Recall that, with the majority of previous works focusing on El Niño, it has been shown that the two types (EP and CP) of each ENSO phase may cause different climatic responses in certain regions, such as reduced rainfall over eastern Australia during EP El Niños [Chiew et al., 1998] contrasted by an increase in precipitation over the same area during CP El Niños [Taschetto and England, 2009].

Most previous studies on El Niño's teleconnective impacts have either applied linear statistical tools, such as correlation analysis [Diaz et al., 2001], or investigated corresponding composites of the climatic observables under study [Hoell et al., 2015]. Both approaches share the limitation that they focus on linear or average interdependencies between ENSO and possible response variables which do not necessarily hold for extreme values. However, with global climate change projected to increase the strength and frequency of both, extreme climatic events [Easterling et al., 2000, Jones et al., 2007, Karl and Trenberth, 2003] as well as extreme ENSO periods [Cai et al., 2014], it has become a pressing issue to specifically assess possible linkages between these two aspects [Allan and Soden, 2008]. Therefore, this chapter is devoted to quantify and spatially resolve signatures of extreme climatic events that are likely to coincide with an ENSO phase of a certain type. In particular, the focus lies here

on heavy precipitation and dry periods, i.e., hydro-meteorological extreme events, as ENSO has been shown to largely affect rainfall patterns at both global and regional scales [Dai and Wigley, 2000, Ropelewski and Halpert, 1987].

The main goal of the study presented in this chapter is to quantify the likelihood of simultaneous or time-delayed occurrences of localized seasonal precipitation extremes along with an ENSO phase of a given type. For this purpose, *event coincidence analysis* (ECA) [Donges et al., 2016, 2011c] has been proven as an appropriate tool for estimating the probability of co-occurrences between such events. This framework has already been successfully applied to quantifying the likelihood of climatic extreme events triggering certain ecological or social responses, such as extreme annual [Rammig et al., 2015] and daily [Siegmund et al., 2016b] tree growth or flowering dates [Siegmund et al., 2016a] as well as the outbreak of epidemics [Donges et al., 2016] or armed conflicts [Schleussner et al., 2016].

In contrast to previously proposed schemes to discriminate El Niño and La Niña into its two different phases, the transitivity index $\mathcal{T}_n^*$ is particularly useful for the analysis of ENSO related climatic extreme responses as it shows distinct peaks (see Fig. 4.4B) and, thus, provides a sharp discrimination compared to previously proposed measures based on average temperature observations [Ashok et al., 2007]. This particular property is important for the present analysis as ECA particularly relies on the assessment of event sequences, which are easily derived for a spike-like signal in contrast to a smoothly varying one. Therefore, $\mathcal{T}_n^*$ is used together with the classical Oceanic Niño Index (ONI) (provided by the Climate Prediction Center of the National Oceanic and Atmospheric Administration, NOAA-CPC) to construct four event series representing EP and CP El Niño and La Niña phases. From there ECA is utilized to quantify simultaneous occurrences of either of the four ENSO types with spatially resolved seasonal precipitation extremes (wet and dry) in boreal fall and winter of the same year as the start of a certain ENSO episode and spring of the following year.

5.2. GPCC rainfall data

As a proxy for extreme rainfall, the gridded monthly precipitation data provided by the Global Precipitation Climatology Centre (GPCC) at a spatial resolution of $2.5° \times 2.5°$ [Schneider et al., 2015] is utilized here. Since reliable estimations of El Niño and La Niña periods according to the ONI and corresponding discriminations into their respective EP and CP flavors by means of $\mathcal{T}_n^*$ are only available for the second half of the 20th century, the analysis is again restricted to the period from 1951 to 2014. The density of stations from which the data is derived varies between 0 and more than 100 per grid cell and for different years [Lorenz and Kunstmann, 2012]. The minimum number of available stations per grid cell during the study period is

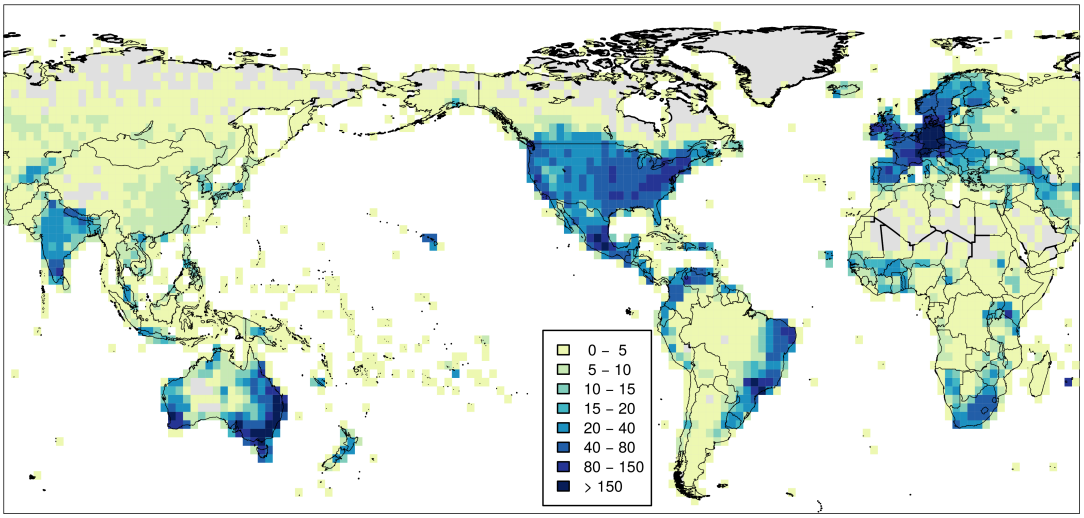

Figure 5.1.: Minimum number of stations per grid cell in the GPCC monthly precipitation data set during the period 1951 to 2014.

displayed in Fig. 5.1. A low density of stations generally results in a lower accuracy and reliability of the data for those areas with only few stations [Rudolf et al., 1994]. Therefore, only grid cells where the minimum number of stations between 1951 and 2014 is larger than 1 are considered, yielding a total number of $N_P = 2248$ cells. From the monthly time series, three separate records for three seasons $\mathcal{S}$ are derived by aggregating the precipitation amounts of the corresponding three-month periods September to November ($\mathcal{S}$ =SON), December to February ($\mathcal{S}$ =DJF) and March to May ($\mathcal{S}$ =MAM). This results in three time series $P_{\mathcal{S},i}(t)$ per grid cell i with $N_t = 64$ annual values each.

5.3. Data preprocessing

Years with seasons $\mathcal{S}$ (DJF, SON, or MAM) exhibiting extraordinary high or low precipitation amounts are defined from the corresponding time series $P_{\mathcal{S},i}(t)$ for each grid cell i individually. Specifically, values above (below) the 80th (20th) percentile $\pi^{+}_{\mathcal{S},i}$ ($\pi^{-}_{\mathcal{S},i}$) in each of the time series $P_{\mathcal{S},i}(t)$ are considered as extraordinary high (low) seasonal precipitation sums. The choice of these particular thresholds ensures the presence of a sufficient number of particularly dry and wet seasons that is comparable with the number of different ENSO phases in the considered time period.

According to these considerations, six binary time series $Y^{\pm}_{\mathcal{S},i}(t)$ per grid cell are obtained as

$$Y^{\pm}_{\mathcal{S},i}(t) = \Theta(\pm P_{\mathcal{S},i}(t) \mp \pi^{\pm}_{\mathcal{S},i}), \tag{5.1}$$

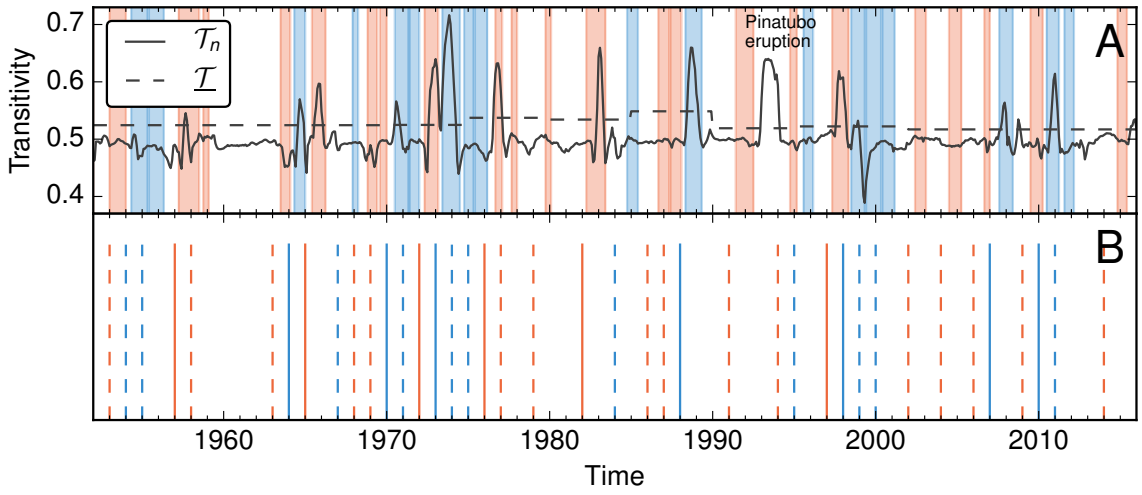

Figure 5.2.: (A) Network transitivity index indicating EP El Niños or La Niñas if its value exceeds the baseline (dashed line) and CP phases otherwise. Red and blue shaded areas indicate periods of El Niño and La Niña according to the ONI. (B) Corresponding series of EP (solid) and CP (dashed) El Niño (red) and La Niña (blue) events, respectively (from Wiedermann et al. [2017b]) .

where $Y^{\pm}_{\mathcal{S},i}(t) = 1$ indicates the presence of a seasonal precipitation extreme at grid cell i during season $\mathcal{S}$ in year t. By following the above procedure, no further deseasonalisation of the precipitation data is necessary, since the grid cell-specific seasonality of precipitation is already taken into account. Furthermore, the events are defined for each location independent and therefore spatial heterogeneity is taken into account as well.

In addition to the precipitation data set, a similar binary indicator time series for the different ENSO phases is obtained from the transitivity index of ENSO flavors $\mathcal{T}^*_n$ (Fig. 5.2). For example an EP El Niño series $Z_{EPEN}(t)$ (solid red line in Fig. 5.2) is obtained with entries $Z(t)_{EPEN} = 1$ if t marks the onset year of an El Niño period (as defined by the ONI) and $\mathcal{T}^*_n$ exceeds $\underline{\mathcal{T}}$ for at least one month of that considered period. All remaining El Niño periods are classified as CP events, yielding a corresponding event series $Z_{CPEN}(t)$ (dashed red line in Fig. 5.2B). The same procedure is applied to La Niña periods, yielding event series $Z_{EPLN}(t)$ and $Z_{CPLN}(t)$, respectively (blue lines in Fig. 5.2B).

5.4. Event coincidence analysis

This work applies event coincidence analysis (ECA), a statistical tool to quantify simultaneities between events in two series [Donges et al., 2016, 2011c, Rammig et al., 2015]. It computes for each grid cell the fraction of EP (CP) El Niño (La Niña) phases that coincide with extreme precipitation sums in SON or DJF of the same

year or MAM of the following year, revealing if the timing of precipitation extremes is non-randomly related to the presence of a given type of ENSO phase. Specifically, the event coincidence rate $ECR^{\pm}_{\mathcal{S},i,\bullet}$ for one pair of ENSO and precipitation event series is given by

$$ECR^{\pm}_{\mathcal{S},i,\bullet} = \frac{\sum_t Z_\bullet(t-\tau) Y^{\pm}_{\mathcal{S},i}(t)}{N_t - \tau}. \tag{5.2}$$

Here, $Z_\bullet(t)$ represents one of the four time series indicating EP and CP types of El Niño and La Niña. The offset τ reads $\tau = 0$ for $\mathcal{S}$ =SON and $\mathcal{S}$ =DJF and $\tau = 1$ for $\mathcal{S}$ =MAM.

To assess the statistical significance of the obtained event coincidence rates, both involved event sequences are assumed to be generated by an underlying Poisson process with resulting events that are thus distributed randomly, independently and uniformly [Donges et al., 2016, 2011c, Rammig et al., 2015, Siegmund et al., 2016a,b]. In particular, it is assumed that the numbers of N_Y precipitation and N_Z ENSO events are sufficiently large and that the events are spaced much more widely than the coincidence interval [Donges et al., 2016], i.e.,

$$1 \ll N_Y \ll N_t \text{ and } 1 \ll N_Z \ll N_t. \tag{5.3}$$

The probability for a specific ENSO or precipitation event to occur randomly at any time in the effective time span $[1, N_t - \tau]$ reads [Donges et al., 2016]

$$p = \frac{1}{N_t - \tau}. \tag{5.4}$$

From this, the probability of a considered precipitation event to coincide with an ENSO event follows as [Donges et al., 2016]

$$p_c(N_Y) = 1 - (1-p)^{N_Y} = 1 - \left(1 - \frac{1}{N_t - \tau}\right)^{N_Y}. \tag{5.5}$$

Here, $(1-p)^{N_Y}$ is the probability that no precipitation event occurs for a given time at which an ENSO event takes place. Based on this expression it is possible to calculate the probability $P(K, N_Z, 1-(1-p)^{N_Y})$ that exactly K coincidences are observed for a given realization of the two Poisson processes from the binomial distribution with N_Z trials and a success probability of $1-(1-p)^{N_Y}$ [Donges et al., 2016]. Thus,

$$P(K, N_Z, p_c(N_Y)) = \binom{N_Z}{K} \left(1 - \left(1 - \frac{1}{N_t - \tau}\right)^{N_Y}\right)^K \left(\left(1 - \frac{1}{N_t - \tau}\right)^{N_Y}\right)^{N_Z - K}. \tag{5.6}$$

A p-value of an observed number of coincidences K_{obs}, i.e., the numerator in Eq. (5.2), with respect to the reference distribution in Eq. (5.6) can be written as the probability to observe K_{obs} or more coincidences by chance [Donges et al., 2016],

$$P(K \geq K_{obs}) = \sum_{K'=K_{obs}}^{N_Z} P(K', N_Z, p_c(N_Y)). \tag{5.7}$$

In the following, an empirical event coincidence rate is considered statistically significant if $P(K \geq K_{obs})$ is smaller than a confidence level of $\alpha = 0.05$. In cases were the requirements in Eq. (5.3) are not fulfilled, it is advisable to use Monte Carlo simulations with appropriately generated pairs of surrogate event series in order to numerically obtain p-values for the observed coincidence rates [Donges et al., 2016, Siegmund et al., 2017]. However, for the present case, the number of N_Y precipitation events is per design, i.e, defining events based on percentiles, sufficiently large but still much lower than the length of the considered time interval. The number of ENSO events varies between $N_{Z_{EPEN}} = 6$ and $N_{Z_{CPEN}} = 16$ and thus also fulfills Eq. (5.3).

5.5. Results

Seasonal precipitation extremes and EP El Niño

First, event coincidences rates between EP El Niños and the timing of seasonal precipitation extremes are investigated. Figures 5.3A,C,E show locations with significant coincidence rates $ECR^{\pm}_{\mathcal{S},i,ENEP}$ between EP El Niños and extremely dry (red squares) and wet periods (blue squares) in SON, DJF and MAM, respectively.

During SON of EP El Niño years, elevated probabilities emerge for extremely dry conditions over Indonesia, the Philippines and the southwestern Pacific islands as well as over northern South America and the northern Amazon Basin (Fig. 5.3A). The latter dry events have been linked to an increased risk of biomass loss in the Amazon which would otherwise serve as a long term carbon sink [Phillips et al., 2009].

For the same season, one also observes an increased likelihood of very wet conditions over the central Pacific islands and the west coast of North America (Fig. 5.3A). Similarly, wet conditions are found over Ecuador and the South American east coast in SON and DJF (Fig. 5.3A,B). Further, extremely wet conditions are discovered over parts of the Chilean Andes in SON (Fig. 5.3A) which may result in an increased risk for the occurrence of floods in this area [Boers et al., 2014a, Bookhagen and Strecker, 2012]. All these observations agree well with previous climatological studies [e.g., Diaz et al., 2001]. Coinciding with EP El Niños, one additionally observes more frequent wet extremes over the Mediterranean region (Fig. 5.3A) in SON, which

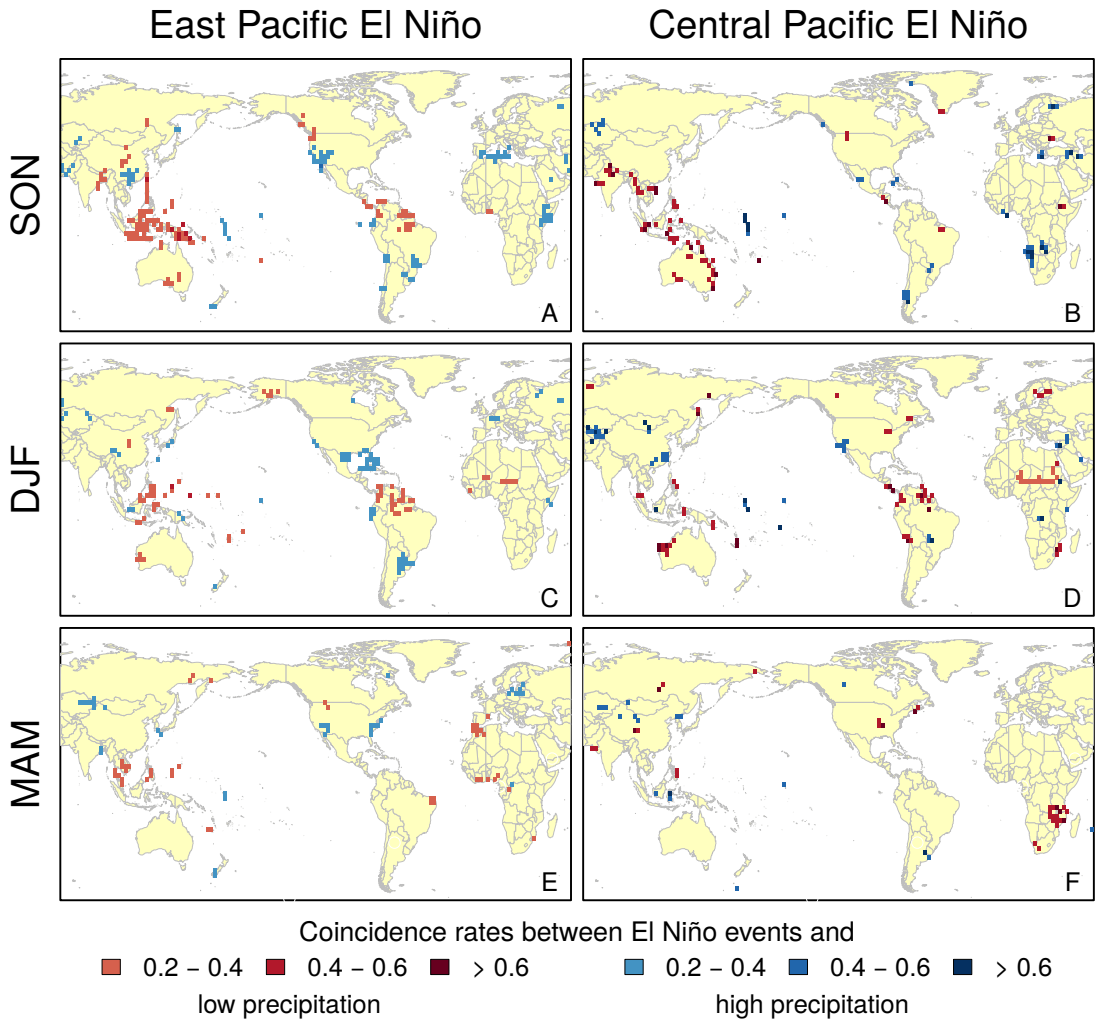

Figure 5.3.: Statistically significant event coincidence rates between years of EP (left column) and CP (right column) El Niños and very dry or very wet seasons (SON, DJF, MAM). Red (blue) squares represent grid cells with significant event coincidence rates between EP or CP El Niño years and seasonal precipitation sums below (above) the 20th (80th) percentile of all years from 1951 to 2014 (from Wiedermann et al. [2017b]).

was partially also reported earlier by Shaman and Tziperman [2010]. Furthermore, the wet conditions over East Africa during the same season (Fig. 5.3A) have previously been described in parts by Camberlin et al. [2001]. The observed dry conditions in southwestern Africa during DJF (Fig. 5.3C) were also reported recently by Hoell et al. [2015]. Finally, significant coincidences between EP El Niños and extremely wet conditions are observed over the Western Caribbean (especially Haiti), which may pose an increased risk for tropical disease outbreaks associated with extreme precipitation [Guzman Herrador et al., 2015].

Generally, the ECA results are in good agreement with previously reported interre-

lations between El Niño and global precipitation, which have mostly been identified using linear correlation analysis. Thus, it is concluded that (i) the application of ECA to assess ENSO-related extreme impacts provides valid results and that (ii) extreme responses to the canonical El Niño display roughly similar spatial patterns as the average statistical interdependency between ENSO-related indices and climatic observables. However, it remains to note that most previous studies have not discriminated between the two El Niño flavors. Thus, the agreement between the recent literature and the results for EP El Niños as presented here suggests that the previously observed linear effects might just be dominated by the (on average stronger) EP events.

Seasonal precipitation extremes and CP El Niño

Next, the focus is shifted to the specific effects of the less intensively studied CP El Niño and event coincidence rates between these phases and seasons with extreme precipitation sums are estimated (Fig. 5.3B,D,E). At first those regions are discussed that display significant coincidence rates for EP El Niños (see Sec. 5.5), but an absence or altered likelihood of seasonal precipitation extremes during CP periods. Notably, the dry events over Central America and the Amazon that frequently occur together with EP El Niños become more likely during DJF (Fig. 5.3D), but show insignificant coincidence rates in SON (Fig,. 5.3B). The latter also holds for the wet events along the western coast of Central and North America that have been observed for EP El Niños. In the same manner, the wet SON patterns over southern China, the Mediterranean and south-east Africa coinciding with EP El Niños disappear during CP events (Fig,. 5.3B). The large-scale dry events over Indonesia observed along with EP events during SON become less spatially coherent for CP El Niños (Fig. 5.3B). For MAM, the wet and dry patterns over Europe become insignificant (Fig. 5.3F). Ultimately, it is observed that the dry pattern in DJF over the central Sahel in Africa observed in coincidence with EP El Niños increases in its spatial extent (Figs. 5.3B,D) and is likely to decrease agricultural production [Epule et al., 2014].

While the above observations indicate decreased or weakened impacts of CP El Niños in comparison with the EP type, additional patterns appear that are not present during EP El Niños but emerge only along with CP phases. Especially dry events become more likely along Australia's east coast during SON (Fig. 5.3B). Such events could thus trigger severe impacts on river ecosystems and agriculture in the area. In particular, dry events in Eastern Australia are likely to cause a cascade of low river inflows, i.e., water scarcity and large scale floodplain forest mortality as well as an increase of toxicity in the surrounding lakes [Leblanc et al., 2012]. These natural hazards in turn have large-scale effects on agricultural production in terms of a severe reduction in irrigated crop yields of up to 99% [van Dijk et al., 2013].

In addition to reduced rainfall responses, significant coincidence rates with wet events arise over the south of Chile pointing towards increased rainfall during CP El Niños as compared to their canonical counterparts. In DJF months coinciding with CP events, new wet patterns over central Asia and China as well as a dry pattern over the north of Chile are emerging (Fig. 5.3D). Ultimately, for MAM one observes a pronounced dry pattern over southeast Africa (Fig. 5.3F). In general, the spatial coherence of seasonal precipitation extremes coinciding with CP El Niños is decreased in comparison with EP phases.

The same analysis is also repeated for extreme precipitation events defined by seasonal sums below the 15th and above the 85th, as well as the 25th and 75th percentile $\pi^{\pm}_{\mathcal{S},i}$, respectively. The corresponding significant coincidences with EP and CP El Niño events are shown in Fig. A.1 and Fig. A.2 in the appendix. As expected, the numerical values of significant coincide rates as well as the spatial extent of the affected regions changes slightly due to the different number of precipitation events N_Y considered in the analysis. However, the key regions that were discussed above are still qualitatively recovered. Therefore, the results presented here are considered sufficiently robust if the quantiles $\pi^{\pm}_{\mathcal{S},i}$ are varied within a reasonable range.

Seasonal precipitation extremes and EP/CP La Niña

Finally, the analysis is repeated for La Niña periods. Various patterns that were reported in earlier studies for the EP, i.e., canonical, La Niña type are recovered (Fig. 5.4A,C,E). Specifically, during SON coinciding with EP La Niña phases (Fig. 5.4A) one finds the expected wet conditions over Australia and Indonesia [Arblaster et al., 2002] and exceptionally dry conditions in parts of southern Europe [Pozo-Vázquez et al., 2005] and the south of Brazil [Ropelewski and Halpert, 1996]. Further, significant coincidence rates with dry events in the middle East and extreme rainfall over central Europe are found. For DJF, the analysis confirms findings by Nicholson and Selato [2000] of wet conditions over southern Africa and dry events over western Africa (Fig. 5.4C), where specifically the latter may cause agricultural losses [Karpouzoglou and Barron, 2014] as well as impose increased health risks [Rataj et al., 2016]. In addition, a prominent extreme precipitation dipole with dry conditions over Mexico and increased precipitation in the southwestern part of Canada is observed. The latter has become an important aspect of local water resource management [Lute and Abatzoglou, 2014], but along with other parameters like wind and temperature also poses the threat of landslides in coastal areas [Guthrie et al., 2010].

For MAM seasons associated with EP La Niñas, one observes a tendency towards extreme rainfall over Amazonia [Rogers, 1988] and parts of northern Australia [Arblaster et al., 2002] (Fig. 5.4). Pronounced significant coincidences between EP La Niñas and very dry conditions in MAM are found over the Caribbean which may negatively impact rainfed agriculture in the corresponding area [Aladenola et al.,

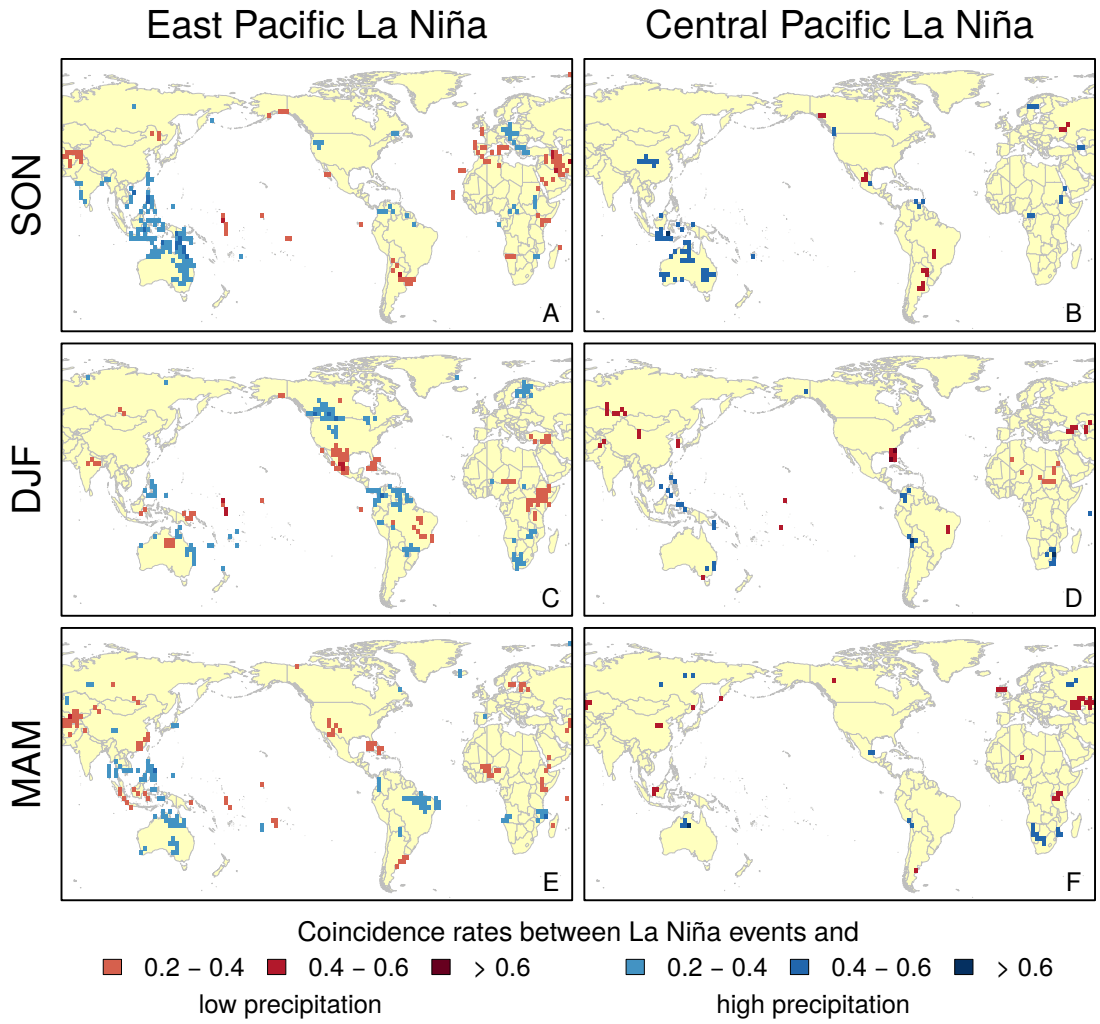

Figure 5.4.: Same as Fig. 5.3 for La Niña periods. (from Wiedermann et al. [2017b])
.

2016].

In contrast to the coherent patterns observed for EP La Niña phases, only a few spatially organized structures are observed along with CP events (Figs. 5.4B, E, F). Most prominently, the previously reported wet conditions over parts of Australia in DJF are recovered [Arblaster et al., 2002, Cai and Cowan, 2009] (Fig. 5.4B). Additionally, strongly reduced rainfall is observed over Florida in DJF and over Scandinavia and the south of Russia in MAM. Thus, similar as for El Niño, it must again be recorded that CP La Niñas coincide with less spatially coherent extreme precipitation responses.

Following upon the presented findings, it is worthwhile to pick up again the recently raised question whether it is actually meaningful to distinguish La Niña into two flavors in the same fashion as El Niño [Kug and Ham, 2011]. While Kao and Yu [2009] and Ashok and Yamagata [2009] advocated for such a distinction, Kug et al.

[2009] and Ren and Jin [2011] argued that based on correlation analyses between La Niña related SST patterns no distinct discrimination into two types is evident. The previous Chapter 4 and the coherent applicability of the transitivity index $\mathcal{T}_n^*$ to La Niña periods already demonstrated that it is indeed meaningful to provide a discrimination of these periods in analogy to El Niño. The results presented in this chapter demonstrate that seasonal precipitation extremes accompanied by EP La Niñas are generally more likely to arise in a spatially coherent way (Fig. 5.4) than such observed along with CP phases. The same property also holds for El Niño periods (Sec. 5.5), which highlights again a symmetry not only within the spatial patterns of El Niño and La Niña themselves (compare the symmetry in sea surface temperature anomalies as presented in Fig. 4.1), but also with respect to their induced extreme precipitation responses. Thus, from an impact-oriented point of view, this work provides further evidence in favor of a distinction between two types of La Niña indicated by the presence or absence of induced seasonal precipitation extremes. In other words, from the viewpoint of extreme events (and thus possibly in contrast to observations based on linear statistical analysis) it seems reasonable to discriminate La Niña into two types in the same way as for El Niño periods.

Again, the analysis is repeated for extreme precipitation events defined by seasonal sums below the 15th (25th) and above the 85th (75th) percentile $\pi^{\pm}_{\mathcal{S},i}$. The corresponding significant coincidences with EP and CP La Niña events are shown in Fig. A.3 and Fig. A.4 in the appendix. Especially for the extremes below (above) the 15th (85th) percentile an expected overall weakening of the significant coincidences' spatial extent is observed. This effect is most likely caused by the decreased number of considered precipitation events N_Y. However, most of the regions discussed above are still recovered and as such the presented results are considered sufficiently robust if the quantiles $\pi^{\pm}_{\mathcal{S},i}$ are varied within a reasonable range.

5.6. Summary

This chapter presented the results of a detailed analysis of ENSO imprints in global patterns of seasonal precipitation extremes. Specifically, El Niño and La Niña were both discriminated into two distinct types (East Pacific and Central Pacific) by utilizing the network transitivity index $\mathcal{T}_n^*$ as derived in Chap. 4. From $\mathcal{T}_n^*$, four distinct event series were defined such that an event in each series corresponds to the presence of the respective type of ENSO event. The globally gridded GPCC rainfall data set was aggregated to grid point-wise seasonal precipitation sums and corresponding event series were obtained by applying the 80th (20th) percentile to define extremely wet (dry) periods. From there, event coincidence analysis (ECA) [Donges et al., 2016, Siegmund et al., 2016b] was applied to identify grid points with significant coincidence rates between different types of ENSO phases and seasonal

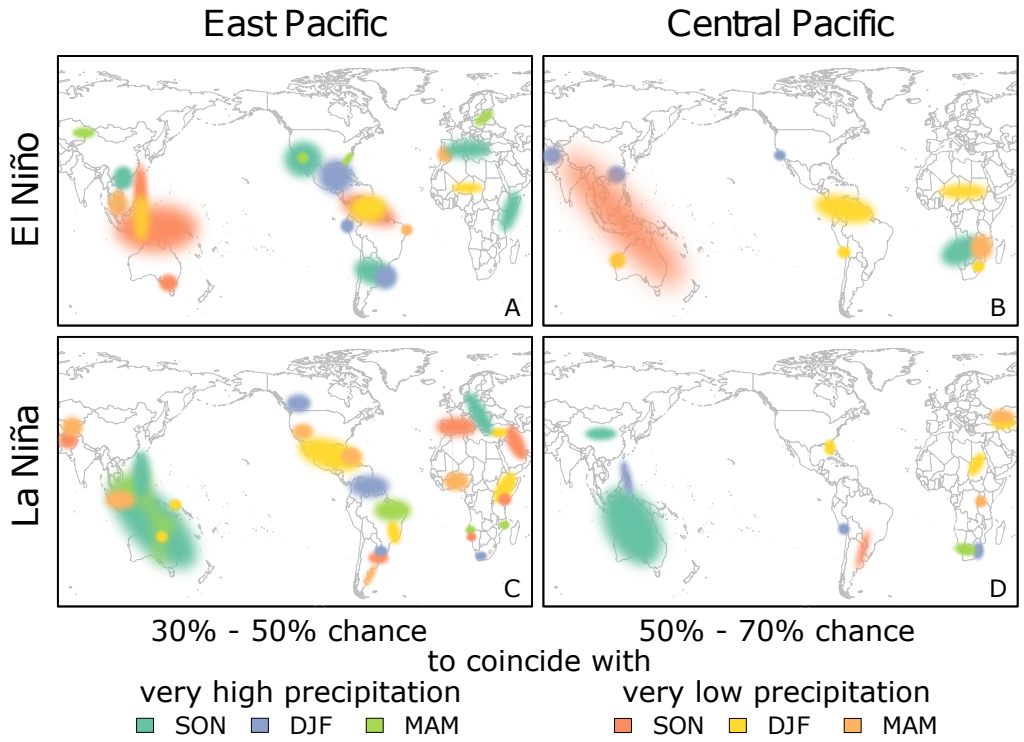

Figure 5.5.: Schematic summary of the results of this work. Shaded areas indicate geographic locations for which extreme (high or low) seasonal precipitation sums exhibit significant coincidence rates with EP (A,C) or CP (B,D) El Niño (A,B) or La Niña phases (C,D). For all indicated areas related to EP (CP) phases, seasonal precipitation extremes occur with a probability of 30-50% (50-70%), respectively (from Wiedermann et al. [2017b]) .

precipitation extremes. Especially the results for the East Pacific El Niño confirmed that previously observed interrelationships based on linear correlation or composite analysis also apply to the associated extremes. In addition, ECA identified further previously unreported impacts of the different ENSO episodes, in particular for the Central Pacific types of both, El Niño and La Niña. This implies that in these cases, standard analysis methods may not be sufficient to capture such interrelations. Moreover, it shows that even though a general linear relationship between ENSO and precipitation might be weak or absent for a given region, dry periods or heavy rainfall can still be possible effects of specific types of ENSO phases. At the same time, it was shown that most previously observed interrelations only hold for events corresponding to East Pacific ENSO flavors. This implies that with a proper discrimination of event types, threats like possible droughts or flooding as well as their socio-ecological consequences could be better anticipated, since for most regions of the Earth, CP types of ENSO events have no or a less spatially organized impact on seasonal precipitation extremes.

In conclusion, this chapter provides a detailed and global overview on the large-scale imprints of different ENSO phases and flavors in extreme seasonal precipitation patterns. All findings are ultimately summarized in Fig. 5.5, which displays schematically all regions for which the four different types of ENSO phases show significant coincidence rates with seasonal precipitation extremes together with the approximate values of these.

Future work should apply the concepts used in this work to also study ENSO related extreme impacts on further climatic observables, such as temperature. Further, a network-based discrimination together with the application of ECA could be applied to directly relate ENSO and its distinct phases more thoroughly to climate impacts, such as those on agricultural yields or water availability. Additionally, an intercomparison between results obtained from linear and event statistics could prove useful in assessing which regions are most affected in terms of extreme climatic responses to the presence of either type and phase of ENSO. Moreover, it could be useful to apply the presented framework also to future climate projections in order to assess possible changes in the spatial extent and frequency of ENSO related extreme events. The analyses presented in Chap. 4 and this chapter also contributed to the ongoing discourse whether or not it is (statistically and climatically) meaningful to discriminate La Niña into two flavors in the same way as done for El Niño. The obtained results suggest that such congruence between ENSO's positive and negative phase is meaningful in both ways as (i) the transitivity index $\mathcal{T}_n^*$ clearly distinguishes between two types of events for El Niño as well as La Niña (Chap. 4) and (ii) event coincidence analysis shows that the EP La Niña displays more spatially coherent seasonal precipitation extremes while fewer impacts are found for its Central Pacific counterpart.

This chapter concludes the assessment of single-layer functional networks and their specific application to classify and characterize different phases of the El Niño Southern Oscillation. It has been shown that methods from network theory provide a valuable complement to existing methods of statistical climatology and serve as a good discriminator between periods in the temporal evolution of a (in this case) climatic system. In the present study, these different periods were characterized by different macroscopic characteristics of the corresponding network structure that were well quantified by an associated global network metric, i.e., the network transitivity $\mathcal{T}^*$.

The next chapter extends these assessments by studying the more general case of an interacting or coupled climate network. In contrast to the focus on a single climatic observable, the specific task will be to investigate the interaction structure between the two major climatic subsystems of the Earth, atmosphere and ocean. Drawing from the observations presented in the previous two chapters, this interaction structure will again be evaluated in terms of local and global network characteristics that

are based on the quantification of transitivity or, in other words, clustering of links between the two corresponding subnetworks.

Chapter 6.

Northern hemispheric atmosphere-ocean interactions

It was shown in the previous chapters that functional climate networks are useful to distinguish different stages of a climatic system as it evolves in time. Specifically, the transitivity $\mathcal{T}^*$, as a global network characteristic that quantifies the underlying climate networks' tendencies to form cluster of nodes with similar corresponding climate dynamics, distinguished well between different phases of the El Niño Southern Oscillation (ENSO). However, this previous study only considered dynamics and associated statistical similarities of anomalies within a single climatic field, e.g., surface air temperature anomalies (SATA). While being considered a variable primarily associated with dynamics in the atmosphere, SATA is highly influenced by dynamics of the ocean, too [Eymard and Reverdin, 2014]. This effect can for example be seen from the high statistical similarity between SATA and sea surface temperature (SST) measurements [Yamasaki et al., 2008]. However, even though being a possible proxy of both, ocean surface and near-surface atmospheric dynamics, the assessment of, e.g., SATA alone certainly does not reveal any possibly interrelations between these two major climatic subsystems of the Earth.

Therefore, and following upon previous studies, this chapter presents an exploratory study to understand and quantify ocean-atmosphere interactions by means of a coupled climate network analysis. Here, one subnetwork represents monthly sea-surface temperature (SST) anomalies, while the other is based on the monthly geopotential height (HGT) of isobaric surfaces at different pressure levels. The study specifically restricts itself to the Northern Hemisphere mid-to-high latitudes during boreal winter at monthly scales. This spatial restrictions has been chosen to complement the recently reported study of ocean-atmosphere interactions in the tropical Pacific [Feng et al., 2012] or over the South Atlantic Convergence Zone [Tirabassi et al., 2015]. The temporal restriction to winter is chosen, since previous studies by means of lagged maximum covariance analysis (MCA) have already revealed that the statistical interrelationship between atmosphere and ocean along the considered domain of interest is strongest and most significant at lags of zero or one month during late

fall and winter [e.g. Czaja and Frankignoul, 1999, Frankignoul and Sennéchael, 2007, Gastineau and Frankignoul, 2015, Liu et al., 2006, Wen et al., 2005].

Section. 6.1 presents the necessary background and gives an overview on recent applications of coupled climate networks. Section 6.2 introduces the data sets that are used in this study. The Sections 6.3 and 6.4 provide the methodological foundation of this work and Sec. 6.5 presents the corresponding results. Ultimately, the findings of the presented study are summarized in Sec. 6.6.

This chapter is primarily based on the study that is presented in P8 [Wiedermann et al., 2016a].

6.1. Climatic and theoretical background

As analogue to the analyses presented in Chap. 4 and 5, most previous studies conducted within the framework of climate networks so far focused solely on the dynamics within a single climatic field or layer. Besides atmospheric characteristics like surface air temperature (see Chap. 4), sea level pressure [Tsonis et al., 2010], or precipitation [Boers et al., 2013, Stolbova et al., 2014], recent studies have also addressed ocean dynamics represented by ocean temperature variability at the surface [Feng and Dijkstra, 2014, Tantet and Dijkstra, 2014] or different depths [van der Mheen et al., 2013] as well as the spatio-temporal variability in the strength of the Atlantic meridional overturning circulation [Feng et al., 2014].

As already stated above it is well known, that the dynamics within the two major subcomponents of the Earth's climate system, ocean and atmosphere, are closely entangled [Frankignoul et al., 2001, Trenberth and Hurrell, 1994]. Examples for these interrelationships include the North Atlantic eddy-driven jet stream [Woollings et al., 2010], or the Pacific Ocean forcing to the atmosphere, which is closely related to the dynamics of ENSO [Wyrtki, 1975]. Further, it has been shown that on time scales of up to one month the ocean is forced by atmospheric circulation, prominently manifested in terms of long-term variability patterns like the Pacific North American pattern [e.g. Frankignoul and Sennéchael, 2007] and the North Atlantic Oscillation [Czaja and Frankignoul, 1999, Gastineau and Frankignoul, 2015].

Inspired by approaches to investigate the interaction structure between different mutually coupled subsystems, such as infrastructure networks [Boccaletti et al., 2014, Buldyrev et al., 2010, Vespignani, 2010], a novel set of coupled network measures has been proposed by Donges et al. [2011b], which provides a general tool set to quantify interdependencies between subcomponents in complex coupled climate networks (see Sec. 2.4 and 3.4). In the course of coupled climate network analysis this framework has been successfully applied to investigate interactions between different layers of geopotential height fields, where each isobaric surface forms a subcomponent of a

larger climate network [Donges et al., 2011b]. Similarly, coupled climate networks have been constructed to study ocean-atmosphere interactions in the tropical Pacific [Feng et al., 2012] or over the South Atlantic Convergence Zone [Tirabassi et al., 2015].

Following upon these previous studies, the work presented in this chapter follows the approach by Donges et al. [2011b] and presents an exploratory study to understand and quantify ocean-atmosphere interactions in the Northern Hemisphere mid-to-high latitudes during boreal winter at monthly scales. To investigate the spatial structure of these complex interaction patterns this work analyzes 18 coupled climate networks consisting of two layers each, one oceanic layer representing sea surface temperature (SST) anomalies and one atmospheric layer representing geopotential height fields (HGT) at different pressure levels from 1000 to 10 mbar covering the entire troposphere as well as the lower stratosphere. As the area of study covers the whole Northern Hemisphere north of 30°N the density of grid points in the considered climate data sets again increase rapidly towards the poles and induces some bias in the unweighted network measures [Radebach et al., 2013, Tsonis et al., 2006]. Therefore, this chapter again relies on the assessment of node splitting invariant (n.s.i.) network characteristic as they are derived in Sec. 2.4. To quantify the topology of coupled climate networks, this work relies on both, local (i.e. node-wise) n.s.i. coupled network measures [Feng et al., 2012, Wiedermann et al., 2013] global network measures [Wiedermann et al., 2016a], which allows to assess and compare the macroscopic correlation structure in each of the 18 coupled climate networks.

First, the n.s.i. cross-degree density of nodes in different subnetworks (Eq. (2.22)) is compared to results obtained from MCA [e.g. Storch and Zwiers, 2001], a well-established standard tool from statistical climatology. As for the strength fields s_i^* in Sec. 4.5, the good qualitative overlap between the two methods confirms the expected similarities [Donges et al., 2015b] and again indicates for the validity of the network approach. Network measures of higher order such as the n.s.i. local cross-clustering coefficient are then utilized to reveal that the statistical interrelation between ocean and atmosphere exhibits a hierarchical structure, in which individual parts or areas of the ocean surface correlate strongly with multiple statistically dissimilar parts of the atmosphere. In that sense, it is shown that methods from coupled (climate) network theory distinguish well between spatial signatures in the two different subnetworks that are related with their mutual interaction. Building upon previous studies by, e.g., Czaja and Frankignoul [1999], Frankignoul and Sennéchael [2007] and Gastineau and Frankignoul [2015] the observed hierarchy can be related to dominant atmospheric patterns forcing the ocean on the time scales investigated in this study. Ultimately, this chapter concludes with an assessment of global interacting network characteristics, i.e, n.s.i. global cross-clustering coefficient (Eq. (2.24)) and n.s.i. cross-transitivity (Eq. (2.26)) which again discriminate well

Layer h	Air pressure p_h [mbar]	Geopotential height Z_h [km]	Threshold $T_h(\rho_h = 0.01)$
0	10	30.9	0.9919
1	20	26.3	0.9936
2	30	23.7	0.9932
3	50	20.5	0.9876
4	70	18.4	0.9781
5	100	16.2	0.9621
6	150	13.7	0.9263
7	200	11.8	0.9166
8	250	10.4	0.8982
9	300	9.2	0.8894
10	400	7.2	0.8895
11	500	5.6	0.8958
12	600	4.2	0.9036
13	700	3.0	0.9119
14	775	2.2	0.9171
15	850	1.4	0.9205
16	925	0.8	0.9215
17	1000	0.1	0.9197

Table 6.1.: Air pressure p_h and associated mean geopotential height Z_h as well as the internal threshold $T_h(d_h = 0.01)$ corresponding to an internal link density of $d_h = 0.01$ for each isobaric surface h.

between parts of the atmosphere, such as troposphere and stratosphere, in terms of topological characteristics related to its interaction with the ocean.

6.2. Data description

The coupled climate networks studied here are constructed from two different climatic observables in order to investigate their interaction structure. One subnetwork is based on monthly anomalies of geopotential height (HGT) fields obtained from the ERA40 reanalysis project of the European Centre for Medium-Range Weather Forecast [Uppala et al., 2005]. The data is given on a regular latitude/longitude grid with a spatial resolution of $\Delta\lambda = \Delta\phi = 2.5°$. Specifically, a total number of 18 layers of HGT fields is investigated. The corresponding pressure values at each isobaric surface as well as the average geopotential height are given in Tab. 6.1. The second subnetwork is constructed from the monthly averaged SST field (HadISST1) provided by the Met Office Hadley Centre [Rayner et al., 2003] with a resolution of $\Delta\lambda = \Delta\phi = 1°$. All grid points with corresponding time series containing missing values are removed from the data set as they represent areas which have been at

least temporarily covered by sea-ice.

The present analysis investigates all grid points north of $\lambda = 30°$N excluding the North Pole itself since due to its latitudinal position at 90°N all nodes i located there would be assigned the trivial node-weight $w_i = 0$. Hence, the corresponding nodes would not contribute meaningfully to any characterization of the resulting network. Both data sets are cropped in their temporal extent to cover the same time span from January 1958 to December 2001 and, hence, each time series consists of $N_t = 528$ temporal sampling points. All above steps of preprocessing yield a total number of $N_{x_s} = 6201$ grid points for the SST data and $N_{x_h} = 3456$ grid points for each isobaric surface h of HGT. For both data sets the annual cycle is removed by subtracting the climatic mean for each month from each time series (as explained in Sec. 3.1). Since the focus lies on the spatial structure of strong statistical interrelationships between ocean and atmosphere during boreal winter months (DJF), the data is restricted to use only the values corresponding to these months which yields a length of each time series of $N_t = 132$ data points.

6.3. Maximum covariance analysis (MCA)

To link the assessment of coupled climate networks with results obtained from classical statistical methods a first part of this study is devoted to the analysis of leading coupled patterns that are obtained from a maximum covariance analysis. To compute coupled patterns, first consider two sets of time series $\{X_{s_n}(t)\}_{n=1}^{N_{xs}}$ and $\{X_{h_m}(t)\}_{m=1}^{N_{x_h}}$ representing the SST field (in what follows indicated by the index s) and one layer h of HGT (see also Tab. 6.1). Further each individual time series in both fields is assumed to be normalized to zero mean and unit variance. The linear lag-zero cross-correlation matrix $\mathbf{C}_{sh}$ with entries $\rho_{s_n h_m}$ is then defined as

$$\rho_{s_n h_m} = \frac{1}{N_t} \sum_{t=1}^{N_t} X_{s_n}(t) X_{h_m}(t), \tag{6.1}$$

where $n = 1, \ldots, N_{x_s}$, $m = 1, \ldots, N_{x_h}$ and N_t denotes the total number of temporal sampling points in the two time series. Due to the heterogeneous spatial distribution of grid points in the present data sets all matrix entries $\rho_{s_n h_m}$ are additionally multiplied by the square roots of the cosine of latitudinal positions $\lambda_\bullet$. This procedure has been shown to ensure equal weighting in regular longitude-latitude grids as the one analyzed here [Buell, 1972, North et al., 1982] and yields the weighted cross-correlation matrix $\mathbf{C}_{sh}^w$ with entries

$$\rho_{s_n h_m}^w = \sqrt{\cos \lambda_{s_n} \cos \lambda_{h_m}} \rho_{s_n h_m}. \tag{6.2}$$

Analogously to empirical orthogonal function (EOF) analysis [e.g. Ghil et al., 2002, Hannachi et al., 2007], MCA identifies orthonormal pairs of coupled patterns $\vec{p}_s^{\,(m)} = (p_{s_1}^{(m)} \dots p_{s_{N_{x_s}}}^{(m)})$ and $\vec{p}_h^{\,(m)} = (p_h^{(m)} \dots p_{h_{N_{x_h}}}^{(m)})$ for $m = 1, \dots, \mathcal{R}$ (with $\mathcal{R}$ being the rank of $\mathbf{C}_{sh}$) which explain as much as possible of the covariance between pairs of time series taken from the two different climatic fields [e.g. Bretherton et al., 1992, Storch and Zwiers, 2001]. The coupled patterns are obtained by solving the singular value problem of the weighted cross-covariance matrix,

$$\mathbf{C}_{sh}^{w} \vec{p}_h^{\,(m)} = \sigma_m \vec{p}_s^{\,(m)}, \tag{6.3}$$

$$\mathbf{C}_{hs}^{w} \vec{p}_s^{\,(m)} = \sigma_m \vec{p}_h^{\,(m)}. \tag{6.4}$$

They are ordered according to their respective singular values σ_k with $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{\mathcal{R}}$. Hence, σ_1 denotes the largest among the $\mathcal{R}$ singular values that can be found to solve the above equations. Therefore, $\vec{p}_h^{\,(1)}$ and $\vec{p}_s^{\,(1)}$ are referred to as the *leading* coupled patterns representing the largest fraction of squared covariance between the two climatic fields given by σ_1^2.

6.4. Coupled network construction and assessment

Coupled climate networks are constructed according to the general construction scheme as presented in Sec. 3.4. Since this study aims to only assess interactions between ocean and atmosphere and does not consider any atmosphere-atmosphere interactions the super-similarity matrix $\mathbf{S}$ as depicted in Eq. 3.7 reduces to

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_s & \mathbf{S}_{s,1} & \cdots & \mathbf{S}_{s,h} & \cdots & \mathbf{S}_{s,18} \\ \mathbf{S}_{1,s} & \mathbf{S}_1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ \mathbf{S}_{h,s} & 0 & \cdots & \mathbf{S}_h & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ \mathbf{S}_{18,s} & 0 & \cdots & 0 & \cdots & \mathbf{S}_{18} \end{pmatrix}. \tag{6.5}$$

Thus, $\mathbf{S}$ is composed of 37 independent block matrices, i.e, the internal similarity matrices $\mathbf{S}_s$ ($N_{x_s} \times N_{x_s}$) and $\mathbf{S}_h$ ($N_{x_h} \times N_{x_h}$) as well as the $N_{x_s} \times N_{x_h}$ cross-similarity matrices between two layers $\mathbf{S}_{sh}$. The entries of all matrices are again given by the absolute value of linear Pearson correlation at zero lag either between pairs of time series from the same (Eq. (3.3)) or different layers (Eq. (3.8)).

Further, it is acknowledged that the statistics related to the quantification of such coupled or interacting networks are so far only defined for the assessment of coupled networks composed of two individual subnetworks. Thus, the one super-similarity matrix $\mathbf{S}$ (Eq. (6.5)) is further reduced to a sequence of 18 similarity matrices $\mathbf{S}_s^h$

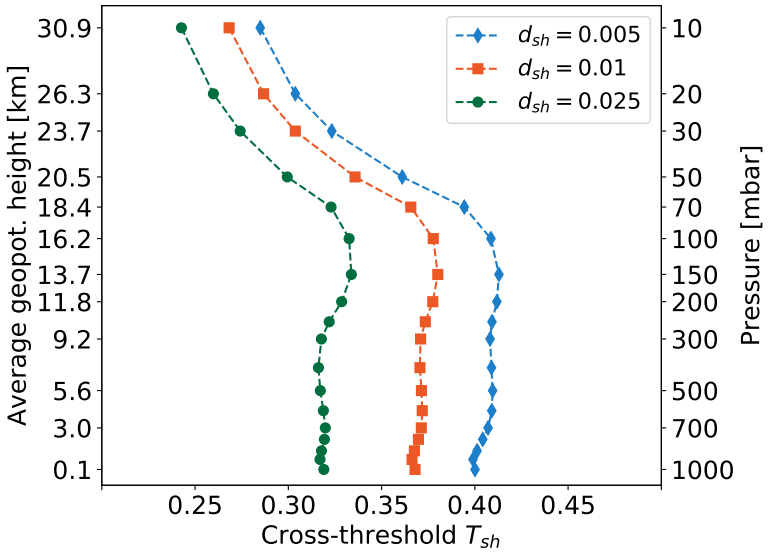

Figure 6.1.: Cross-threshold T_{sh} between the subnetwork constructed from the SST field and all 18 isobaric surfaces of HGT in winter for different standard (unweighted) cross-link densities (adapted from Wiedermann et al. [2016a]).

between the SST field and one individual HGT layer h,

$$\mathbf{S}_s^h = \begin{pmatrix} \mathbf{S}_s & \mathbf{S}_{s,h} \\ \mathbf{S}_{h,s} & \mathbf{S}_h \end{pmatrix}. \tag{6.6}$$

For each pair of SST and HGT layers the (extended) adjacency matrix $\mathbf{A}_s^{h+}$ is then obtained by individually thresholding each of 37 block matrices that constitute the super-similarity matrix $\mathbf{S}$ as well as the respective reduced variants $\mathbf{S}_s^h$. Recall, that according to Sec. 3.3 and 3.4 internal (T_s and T_h) and cross-thresholds T_{sh} are obtained by prescribing fixed individual internal (d_s and d_h) and cross-link densities d_{sh} (Eq. (2.18)). In this work the desired internal link densities are fixed to $d_s = d_h = 0.01$ meaning that only nodes with an absolute correlation above the empirical 99th percentile of absolute correlations between all time series within each field are connected. This condition then leads to internal correlation thresholds $T_s = 0.8101$ for the SST field and a separate T_h for each isobaric surface of HGT (Tab. 6.1). Usually, the dynamics within the different climatic fields show much higher cross-correlations than between them, e.g., due to the stratification of the atmosphere [Donges et al., 2011b]. This fact is accounted for by assuming the fraction of significant interactions between the climatic fields to be lower than those within them. Specifically, a cross-link density of $d_{sh} = 0.005 < d_s = d_h$ is chosen, which is lower than the internal ones. This yields a cross-threshold T_{sh} for each layer of HGT individually (Fig. 6.1). All internal thresholds T_s and T_h are significantly larger than

the obtained cross-thresholds T_{sh}. Thus, setting a global link density or threshold would cause no or few cross-links to be present between the two fields or respective subnetworks. It is noted that all links in the networks represent correlations that are significant at least at the 95% confidence level of a standard t-test, where the degrees of freedom are determined by the total number of temporal sampling points N_t in each of the time series ($N_t = 132$ thus yields 130 degrees of freedom when neglecting the presence of serial correlations in the individual time series).

The different values of T_{sh} already give an impression of the strength of correlation between the SST field and the different isobaric layers: low thresholds generally indicate weaker correlations while high thresholds imply stronger similarity between both fields. Further, it is noted that the resulting cross-thresholds vary smoothly with the choice of cross-link density (Fig. 6.1). Thus the construction mechanism is considered to be sufficiently robust to the actual choice of cross-link density.

Using the different thresholds introduced above, the coupled climate network's (extended) adjacency matrices $\mathbf{A}_s^{h+}$ can be obtained by individually thresholding the absolute correlation values in the block matrices of $\mathbf{S}_s^h$,

$$\mathbf{A}_s^{h+} = \begin{pmatrix} \Theta(\mathbf{S}_s - T_s) & \Theta(\mathbf{S}_{s,h} - T_{sh}) \\ \Theta(\mathbf{S}_{h,s} - T_{sh}) & \Theta(\mathbf{S}_h - T_h) \end{pmatrix}.$$

Note again, that the climate networks in this study are constructed from climate data sets where the density of grid points and, hence, the density of nodes in the network, rapidly increases towards the North Pole. In order to avoid a bias in the evaluation of the climate network's structure, this effect is again accounted for by relying on node-weighted network measures where nodes i are valued with a gradually decreasing weight according to their latitudinal position λ_i as one moves from the equator towards the pole (Sec. 3.1), $w_i = \cos \lambda_i$.

To quantify the correlation structure between ocean and atmosphere at each node this work focuses on two previously defined node weighted local network measures, the n.s.i cross-degree density κ_i^{m*} as given in Eq. (2.22) [Feng et al., 2012] and the n.s.i. local cross-clustering coefficient $\mathcal{C}_i^{m*}$ as given in Eq. (2.20) [Wiedermann et al., 2013]. Donges et al. [2015b] showed that for the unweighted case cross-degree (density) and leading coupled patterns display strong similarity if the first coupled patterns explain a high fraction of the system's covariance. A similar assessment can be made for the similarity between the leading coupled patterns obtained from a weighted cross-covariance matrix and the n.s.i. cross-degree or n.s.i. cross-degree density (see Appendix Sec. A.2).

As seen from Chap. 4, network characteristics based on the counting of triangular linkages between nodes, i.e., strongly connected clusters within the network, seem to provide useful statistics that allow for a discrimination of a sequence of networks under study into different types or classes. In order to discriminate ocean-atmosphere

interactions with respect to the different atmospheric layers in a similar fashion, this work relies on versions of the clustering coefficient and transitivity that are generalized to the node weighted cases of interacting network characteristics. Specifically, this work employs the n.s.i. global cross-clustering coefficient $\mathcal{C}^*_{nm}$ (Eq. (2.24)) and the n.s.i. cross-transitivity $\mathcal{T}^*_{nm}$ (Eq. (2.26)).

6.5. Results

Maximum covariance analysis (MCA)

The analysis starts by computing the leading coupled patterns between the SST field and the 18 HGT layers for boreal winter (DJF). Figure 6.2 displays the results for three representative layers of HGT at 50 mbar, 100 mbar and 500 mbar.

By applying MCA, one detects coherent large-scale patterns of winter SST, which co-vary with the winter atmospheric circulation structures instantaneously. The leading MCA patterns explain rather large amounts of 42%, 63% and 70% (for the 500, 100 and 50 mbar pressure level, respectively) of the squared covariance. At all levels, the leading MCA mode displays significant SST anomalies over the North Pacific with maximum values along the sub-Arctic front near 40°N and anomalies of the opposite sign along the western coast of North America (Fig. 6.2A,C,E) [An and Wang, 2005, Frankignoul and Sennéchael, 2007]. Over the North Atlantic, a dipole structure is seen between the northern part of the Gulf Stream and the Atlantic Ocean south of Greenland including parts of the Davis Strait and the North Atlantic current. This pattern resembles the first SST EOF for the Northern Hemisphere during boreal winter [Wiedermann et al., 2016a].

This general SST pattern is co-varying with a pressure anomaly pattern showing a hemispheric annular-like structure in the upper troposphere and lower stratosphere (Fig. 6.2B,D). In the mid-troposphere (Fig. 6.2F), this pattern displays wave-like deviations from the annular structure, which show distinct similarities with the wave-train structure of the Pacific North American (PNA) pattern. Therefore, the leading MCA mode relates negative SST anomalies along the sub-Arctic front with a positive PNA phase.

Beyond the results presented here, several studies applied lagged MCA between SST and mid-tropospheric circulation fields, and showed that the squared covariance fraction is strongest and most significant at lags of 0 and 1 month during late fall and winter [e.g. Czaja and Frankignoul, 1999, Frankignoul and Sennéchael, 2007, Gastineau and Frankignoul, 2015, Liu et al., 2006, Wen et al., 2005]. This points to the forcing of the SST by the dominant atmospheric pattern, which is the PNA pattern over the Pacific-North American sector [e.g. Frankignoul and Sennéchael, 2007] and the North Atlantic Oscillation (NAO) over the North Atlantic-European

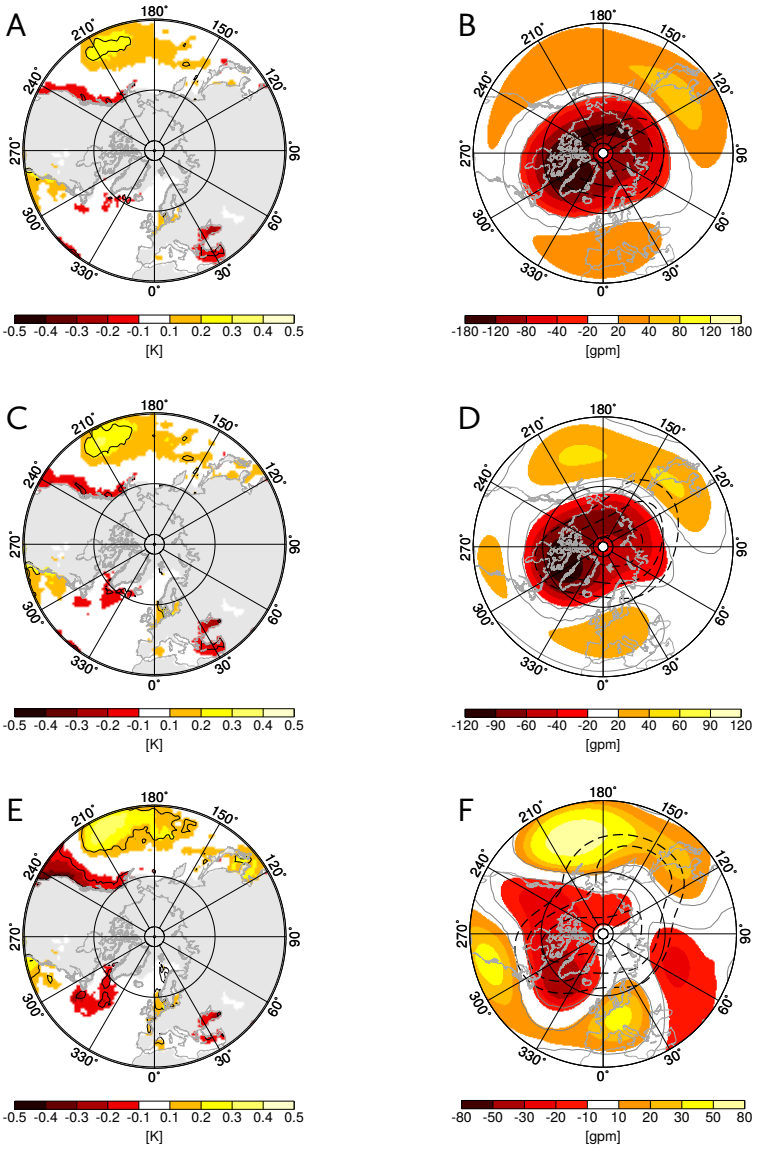

Figure 6.2.: Leading coupled patterns obtained from MCA between the SST field and three layers of geopotential height at 50 mbar (A and B), 100 mbar (C and D) and 500 mbar (E and F) in winter (DJF). The left column (A, C and E) displays the component in the SST and the right column (B, D and F) that in the respective HGT field. All spatial patterns are shown as regression maps obtained by regressing SST anomalies and geopotential height anomalies onto associated time series for the geopotential height field for the respective MCA mode. Statistically significant areas at the 95% confidence level based on a two-tailed Student's t-test are shown as black contours (for the SST maps) and gray contours (for the geopotential height maps). Dashed lines indicate the climatic mean geopotential height fields (from Wiedermann et al. [2016a]).

region [Czaja and Frankignoul, 1999, Gastineau and Frankignoul, 2015]. On the other hand, results of lagged MCA analyses with the ocean leading by 1 to 4 months in Frankignoul and Sennéchael [2007] and Gastineau and Frankignoul [2015] suggest that the SST anomalies have a substantial influence on the large-scale atmospheric circulation at these time-scales.

Local coupled network measures

In order to first demonstrate the general consistency of coupled climate network analysis in comparison with MCA, the analysis continues by generating coupled climate networks between the SST field and the three previously considered layers of geopotential height (500 mbar, 100 mbar, 50 mbar). The n.s.i. cross-degree densities κ_i^{h*} and κ_i^{s*} are expected to display similar spatial structures as the corresponding leading coupled patterns [Donges et al., 2015b] since the latter explain a high share of the cross-covariance between both fields (see Appendix Sec. A.2).

As demonstrated in Fig. 6.3, the results for κ_i^{s*} and κ_i^{h*} indeed match well the results obtained from the MCA when comparing the locations of maximum values in the coupled network's n.s.i. cross-degree densities to those of maximum or minimum values in the leading mode of the MCA. Note, that the n.s.i. cross-degree densities κ_i^{s*} and κ_i^{h*} take, per definition, only positive values, while coupled patterns display both, positive and negative values. Hence, κ_i^{s*} and κ_i^{s*} only reproduce structures that coincide with the absolute values of the leading coupled patterns. However, as only a certain percentage of squared covariance is explained by the leading coupled patterns, one also notes differences between the patterns revealed by the two methods. In particular, the negative center of action around the North Pole that is detected by MCA is only weakly present in the cross-degree density fields κ_i^{s*} for the 50 and 100 mbar HGT fields (compare Fig. 6.2B,D with Fig. 6.3B,D). For the ocean, preferably the marked structures in the leading coupled patterns in both the Atlantic and the Pacific are well recovered by the cross-degree density κ_i^{h*} while some of the weaker structures, e.g. in the Black Sea, are missing.

Network analysis, however, allows to undertake a further in-depth analysis of the correlation structure between the different layers beyond the information provided by MCA. The n.s.i. local cross-clustering coefficients $\mathcal{C}_i^{h*}$ and $\mathcal{C}_i^{s*}$ (Eq. (2.20)) give the probabilities that the dynamics at a grid point in, e.g., the SST field is similar to that at two grid points in the HGT field, which behave themselves statistically similar. Figure 6.4 presents the results for the n.s.i. local cross-clustering coefficients $\mathcal{C}_i^{h*}$ for nodes in the SST field (Fig. 6.4A,C,E) and $\mathcal{C}_i^{s*}$ for nodes in the HGT fields (Fig. 6.4B,D,F). Most nodes in the SST field tend to display a low n.s.i. local cross-clustering coefficient $\mathcal{C}_i^{h*} < 0.2$ (Fig. 6.4A,C,E) and, thus, preferentially correlate with nodes in the HGT fields that are mutually dissimilar and therefore disconnected (Fig. 6.5). In contrast, many nodes in the HGT fields exhibit a comparatively high

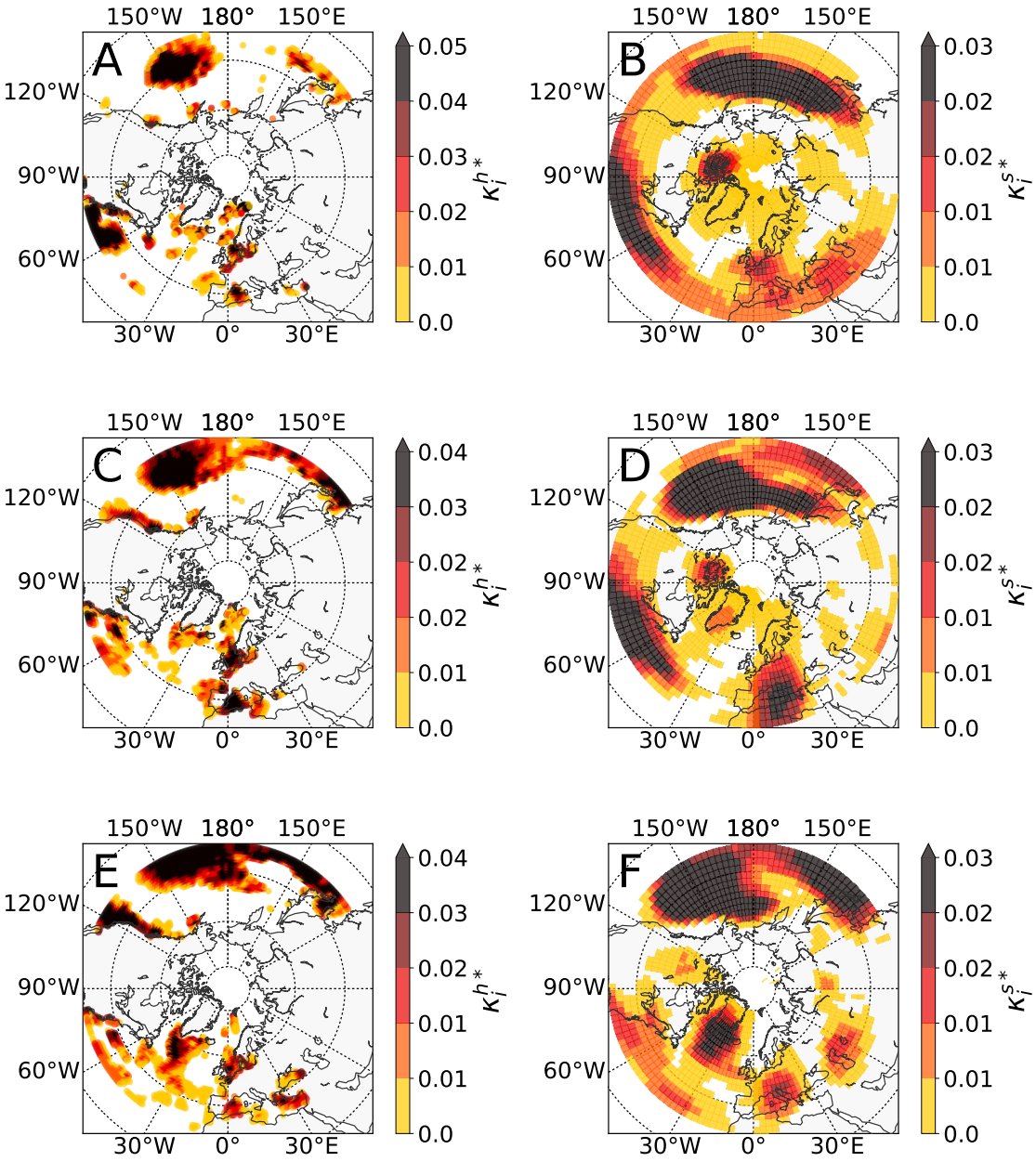

Figure 6.3.: N.s.i. cross-degree density for coupled climate networks constructed from the SST field and three layers of geopotential height at 50 mbar (A and B), 100 mbar (C and D) and 500 mbar (E and F) for winter months (DJF). The left column (A, C and E) displays the n.s.i. cross degree density κ_i^{h*} for links pointing from the SST into the HGT subnetwork while the right column (B, D and F) displays the n.s.i. cross-degree density κ_i^{s*} for links pointing from the HGT into the SST subnetwork. Only nodes with $\kappa_i^{h*} > 0$ and $\kappa_i^{s*} > 0$ are shown (adapted from Wiedermann et al. [2016a]).

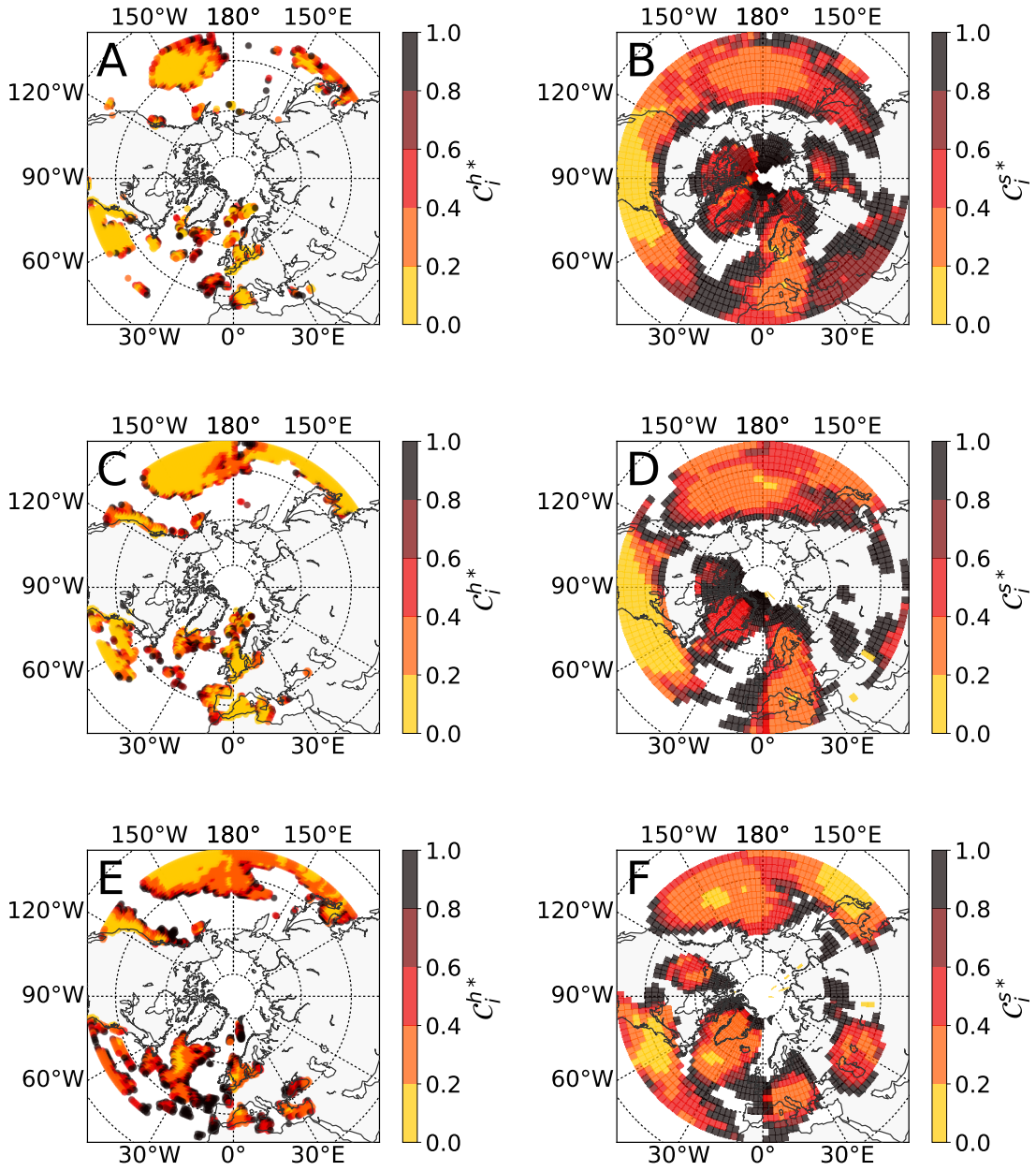

Figure 6.4.: As in Fig. 6.3 for the n.s.i. local cross-clustering coefficients $\mathcal{C}_i^{h*}$ and $\mathcal{C}_i^{s*}$ (adapted from Wiedermann et al. [2016a]).

or intermediate n.s.i. local cross-clustering coefficient $0.4 < \mathcal{C}_i^{s*} < 1$ (for one of the most prominent examples compare nodes located at or above the Pacific in Fig, 6.4C,D). Quantitatively, for the combination of the SST and the 500 mbar HGT field one finds an n.s.i. global cross-clustering coefficient (Eq. (2.24)) of $\mathcal{C}_{sh}^* = 0.16$ for SST nodes and $\mathcal{C}_{hs}^* = 0.28$ for 500 mbar HGT nodes. Ignoring those nodes in the

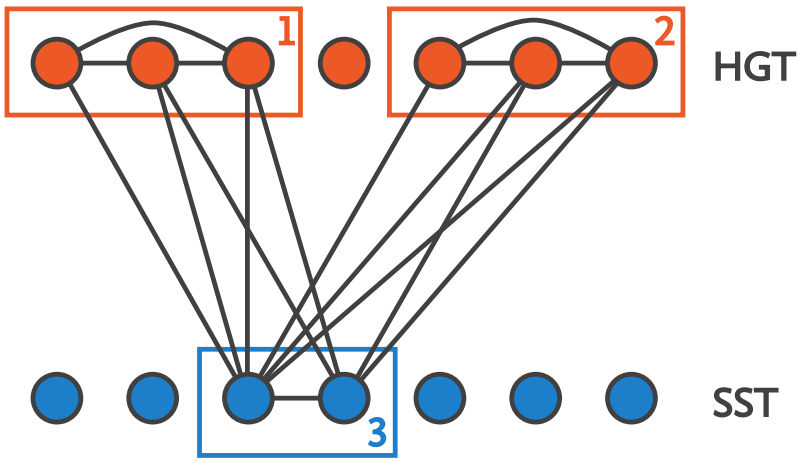

Figure 6.5.: Schematic explanation of the observed quantitative differences in the n.s.i. local cross-clustering coefficients for nodes in the SST and HGT fields. Nodes in the ocean (box 3) tend to connect with statistically dissimilar and thus unconnected clusters of nodes in the atmosphere (such as nodes in box 1 and 2). Hence, the n.s.i. local cross-clustering coefficient $\mathcal{C}_i^{h*}$ only takes low values. In contrast, nodes in the atmosphere, e.g. from box 1, likely connect with clusters in the SST field, such as nodes exclusively in box 3. This results in a high n.s.i. cross local-clustering coefficient $\mathcal{C}_i^{s*}$ for nodes in the atmosphere (from Wiedermann et al. [2016a]).

averaging that display zero n.s.i. cross-degree density yields values of ${\mathcal{C}_{sh}^*}' = 0.42$ and ${\mathcal{C}_{hs}^*}' = 0.52$ (note that this definition is different from the one presented in Eq. (2.24) the contribution of nodes with no links to the opposite subnetwork is specifically excluded). The n.s.i. cross-transitivity (Eq. (2.26)) which weighs nodes according to their n.s.i. cross-degree density gives values of $\mathcal{T}_{sh}^* = 0.2$ and $\mathcal{T}_{hs}^* = 0.25$ for nodes in ocean and atmosphere, respectively. For all three measures, the values computed for the atmospheric subnetwork exceed those for the ocean and thus consistently imply that nodes in the ocean are less likely to connect with mutually connected nodes in the atmosphere than vice versa.

To further quantify the asymmetries in the correlation structure between ocean and atmosphere, Fig. 6.6 investigates for each node with a given n.s.i. cross-degree density its corresponding n.s.i. local cross-clustering coefficient in a coupled climate network composed of the SST and 500 mbar HGT fields. This layer is chosen as it provides a good indication of the atmospheric circulation over the area of study [Gastineau and Frankignoul, 2015, Kushnir et al., 2002]. Furthermore, it displays among the highest values of T_{sh} according to Fig. 6.1, which has similarly been described as a *strong statistical signal* by Frankignoul and Sennéchael [2007].

For nodes in the SST field (Fig. 6.6A), $\mathcal{C}_i^{h*}(\kappa_i^{h*})$ tends to follow a power-law, $\mathcal{C}_i^{h*} \sim (\kappa_i^{h*})^{-\alpha}$, which indicates a hierarchical network structure [Ravasz et al., 2002, Ravasz and Barabási, 2003] that, in contrast, is absent for nodes in the HGT field

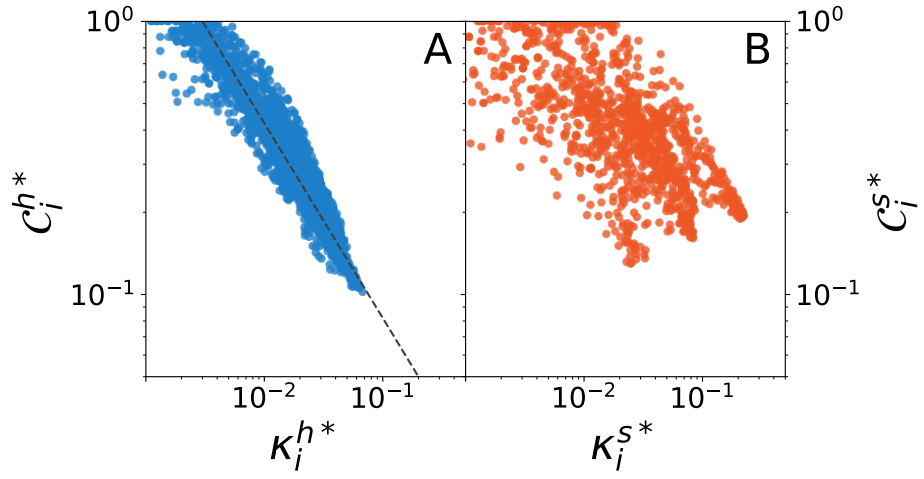

Figure 6.6.: N.s.i. local cross-clustering coefficients $\mathcal{C}_i^{h*}(\kappa_i^{h*})$ for nodes in the SST field (A) and $\mathcal{C}_i^{s*}(\kappa_i^{s*})$ for nodes in the 500 mbar HGT field (B) as functions of the respective n.s.i. cross-degree densities. The dashed line in (A) indicates the relationship $\mathcal{C}_i^{h*} \sim (\kappa_i^{h*})^{-\alpha}$ (here with $\alpha = 0.94$) expected for traditional network measures $C_v(k_v)$ in the case of hierarchical network structures [Ravasz et al., 2002, Ravasz and Barabási, 2003] (adapted from Wiedermann et al. [2016a]).

(Fig. 6.6B). Here, the term *hierarchical* implies that nodes in the SST field strongly correlate with disconnected clusters of statistically similar nodes in the HGT field as depicted in Fig. 6.5. This deduction is further supported by the fact that for the HGT field, the distribution of combinations of $\mathcal{C}_i^{s*}$ and κ_i^{s*} is more widely spread and $\mathcal{C}_i^{s*}$ generally takes higher values than $\mathcal{C}_i^{h*}$. This implies that nodes in the HGT field show a stronger tendency to correlate with mutually connected nodes in the SST field, which can be assumed to display a strong statistical similarity among themselves [Molkenthin et al., 2014, Tupikina et al., 2014]. To test for the robustness of these results the same analysis as presented in Fig. 6.6 is carried out for internal link densities of $\rho_s = \rho_h = 0.02$ and $\rho_s = \rho_h = 0.05$ and corresponding cross-link densities $\rho_{sh} = 0.01$ and $\rho_{sh} = 0.025$ (see Figs. A.5 and A.6 in the appendix). Even though the power-law exponent α slightly decreases towards zero with increasing link densities, the qualitative findings remain unchanged and, hence, the analysis is again considered to be sufficiently robust with respect to the actual choice of link densities.

As a remark it is noted, that a general tendency to display comparatively low values of n.s.i. cross-degree density and increased values of n.s.i. local cross-clustering coefficient is observed for nodes at the boundaries of a cluster that links with the opposite field (Fig. 6.3 and Fig. 6.4). In contrast, nodes located towards the center of these clusters display increased n.s.i. cross-degree density, which is in general to be expected from the continuity of the underlying system. However, in that case one also

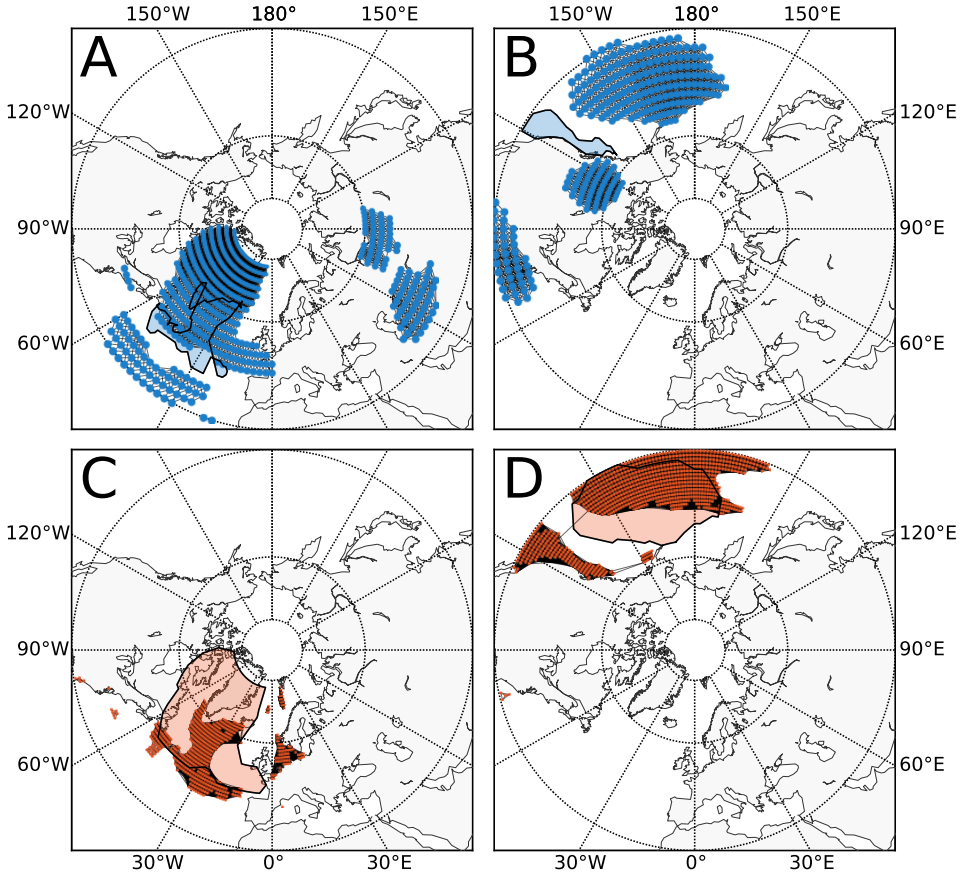

Figure 6.7.: Visualization of a selection of nodes that are relevant for the observed hierarchical network structure. (A,B) Two clusters of nodes in the SST field (blue and orange shaded polygons) that show positive values of κ_i^{h*} with the 500 mbar HGT field (compare Fig. 6.3E). Correspondingly, colored scatter points denote nodes in the HGT field that the considered SST nodes are connected with. All links that mutually connect the resulting HGT nodes are displayed as well. (C,D) The same for the two largest patches of nodes in the 500 mbar HGT field that where detected in (A,B). Colored scatter points now indicate all nodes in the SST field that are connected with these patches (from Wiedermann et al. [2016a]).

notes tendencies for decreased values of n.s.i. local cross-clustering coefficients. This observation is a result of the fact that especially those nodes at the boundary with only one link to the opposite field show by definition the highest value of the n.s.i. local cross-clustering coefficient, $\mathcal{C}_i^{s*} = 1$ or $\mathcal{C}_i^{h*} = 1$. With increasing n.s.i. cross-degree density this measure converges to a more reasonable estimate of a node's tendency to cluster.

One way to address this issue in the future would be to subtract the squared sum of weights w_j of all neighbors j of a considered node i from the numerator in Eq. (2.20). Such procedure would, however, introduce a non-standard network

measure whose properties should be assessed thoroughly in future research before applying it to climatic studies. To this end, it is acknowledged that the concerned nodes do not play a crucial role for the propositions put forward in this section, since they are ultimately dealt with in the following subsection by the assessment of n.s.i. cross-transitivity which weighs those corresponding nodes much lower than those with a high n.s.i. cross-degree density. Furthermore, the corresponding nodes only manifest in the very upper left corners of Fig. 6.6A,B. In that case they do not contribute significantly to the observed relationship between n.s.i. cross-degree density and n.s.i. local cross-clustering coefficient and have no further impact on the qualitative statements put forward above.

Following upon the quantitatively observed hierarchy, Fig. 6.7 allows for a visual inspection of some illustrative parts of the corresponding network structure. In particular, two selected patches of nodes in the SST field that show high values of κ_i^{h*} with the HGT field are displayed (blue shaded polygons in Fig. 6.7A,B) together with their corresponding neighboring nodes in the HGT field as well as all links between those nodes (respective blue scatter in Fig. 6.7A,B). While ignoring very small clusters, one finds a total number of four (three) substantial mutually disconnected patches of nodes in the HGT field that correlate with the respective ocean patches. Vice versa, by selecting the resulting two largest patches of nodes in the HGT field above both oceans (orange shaded polygons in Fig. 6.7C,D) it is observed that each of the patches only correlates with two disconnected patches of nodes in the SST field that are of a relevant spatial extent to have an effect on the estimation of $\mathcal{C}_i^{s*}$. Thus, the resulting n.s.i. cross-clustering coefficient $\mathcal{C}_i^{s*}$ for nodes in the atmosphere is expected to be exceed $\mathcal{C}_i^{h*}$ for nodes in the ocean as the latter correlates with more mutually disconnected clusters of nodes than vice versa.

Comparing the observed node patches in the HGT field with atmospheric patterns of large-scale variability patterns [Handorf and Dethloff, 2012], the two atmospheric clusters in the HGT field that are located above the Atlantic (blue scatter in Fig. 6.7A,B) may be related with the NAO. Correspondingly, the three patches located above the Pacific (blue scatter in Fig. 6.7B) coincide well with the spatial signature of the PNA pattern. Taking into account past studies that applied lagged correlation analysis one notes that on the time scales considered in this study the atmosphere serves as a driving force of the ocean along the spatial domain that is of interest here [e.g. Czaja and Frankignoul, 1999, Frankignoul and Sennéchael, 2007, Gastineau and Frankignoul, 2015]. Thus, the hierarchical network structure might be a result of this aforementioned atmospheric forcing to the ocean. Furthermore, the framework of coupled climate networks and the methodology put forward in this chapter resolves the corresponding induced correlation structure between the two climatic subsystems in a spatially explicit way (Fig. 6.7), such that it enables to specifically detect forcing and forced areas in atmosphere and ocean, respectively.

Choosing different HGT layers up to 200 mbar yields very similar results (see Fig. A.7 in the appendix). This aligns well with previous results by Czaja and Frankignoul [1999] and Frankignoul and Sennéchael [2007] who observed comparable spatial statistical patterns at each tropospheric level. Following the above lines of thought, future work should investigate coupled climate networks constructed from lagged cross-correlations to investigate whether the observed structures are indeed a result of short-term atmosphere-to-ocean forcing. Such procedure would, however, require the derivation of novel directed interacting network measures, which in turn would provide a valuable extension to the framework of climate network analysis.

Global measures

So far, this study focused on three atmospheric layers, namely the 50 mbar, 100 mbar and 500 mbar HGT field. Specifically for the latter case, a further in-depth analysis of the observed hierarchical structures was carried out by means of assessing the power-law relationship between κ_i^{h*} and $\mathcal{C}_i^{h*}$ as well as investigating the distinct spatial distribution of nodes and links in ocean and atmosphere that obey the observed hierarchical organization (Fig. 6.7). To show that these structures are (i) not only present for the 500 mbar field and (ii) their observations are robust with respect to the choice of link densities this section now investigates global network characteristics that provide a macroscopic description of the observed network structures.

Specifically, the n.s.i. cross-transitivity $\mathcal{T}_{sh}^*$ is computed for nodes in the SST field and $\mathcal{T}_{hs}^*$ is computed for nodes in each of the HGT fields according to Eq. (2.26) together with the n.s.i. global cross-clustering coefficients $\mathcal{C}_{sh}^*$ and $\mathcal{C}_{hs}^*$, respectively (Eq. (2.24)). Note again that the latter are defined as the weighted means of their local counterparts that are presented in Fig. 6.4, where nodes with no links to the opposite field are weighted in the same fashion as those with adjacent cross-links. In contrast, the n.s.i. cross-transitivity assigns nodes a weight corresponding to their n.s.i. cross-degree (which for the case of no adjacent cross-links takes a value of zero) and, thus, excludes them from the averaging.

The corresponding results are summarized in Fig. 6.8. Both characteristics, $\mathcal{T}_{sh}^*$ and $\mathcal{C}_{sh}^*$, show their maximum values at around 10 km altitude (250 mbar) (Figs. 6.8A and 6.8B). For the same quantities, distinct minima at 850 mbar (1.4 km) coincide with the transition from the atmospheric boundary layer to the lower troposphere as also found in Donges et al. [2011b]. For all layers above 100 mbar, $\mathcal{T}_{sh}^*$ remains almost constant at low values. Hence, $\mathcal{T}_{sh}^*$ and $\mathcal{C}_{sh}^*$ seem to naturally discriminate between three different atmospheric layers: below 850 mbar (atmospheric boundary layer), between 850 mbar and 100 mbar (free troposphere) and above 100 mbar (lower stratosphere).

For the global measures computed over all nodes in the HGT field, the n.s.i. cross-transitivity $\mathcal{T}_{hs}^*$ shows almost constant values for all layers below 200 mbar and,

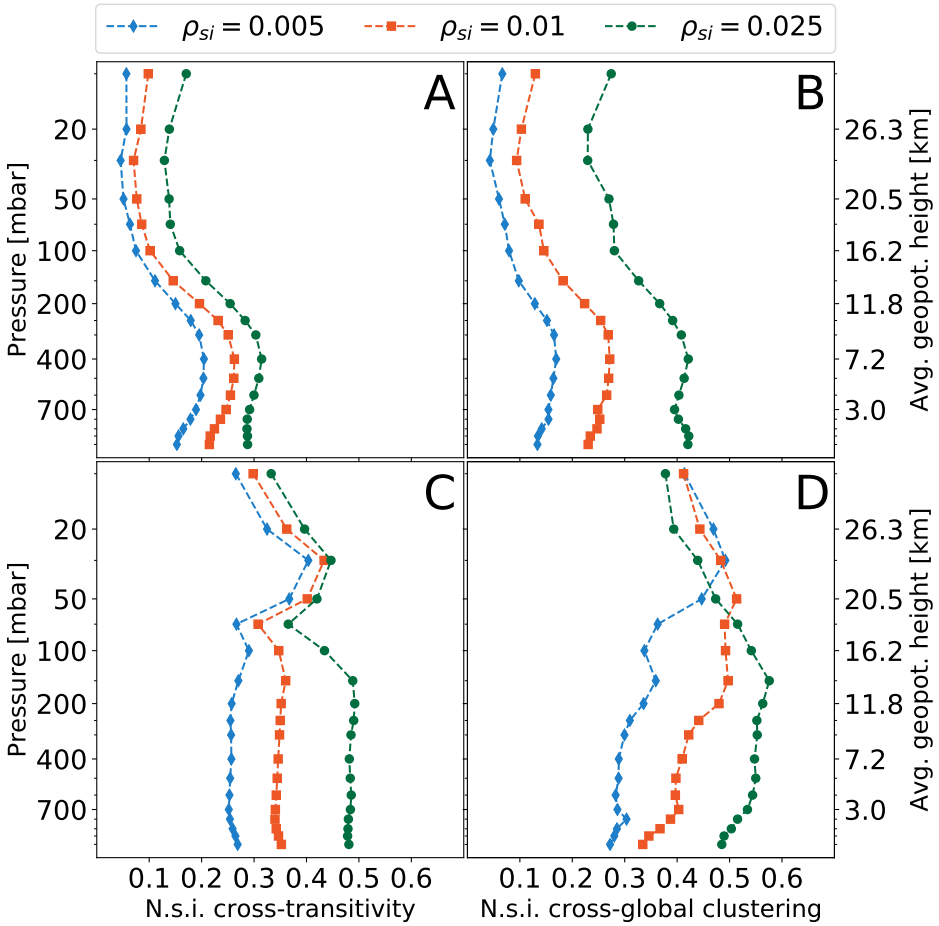

Figure 6.8.: Global coupled network measures computed for all 18 coupled climate networks: (A) N.s.i. cross-transitivity and (B) n.s.i. global cross-clustering coefficient (B) taken over all nodes in the SST field. (C) and (D) display the respective measures computed over all nodes in the HGT field. To demonstrate the robustness and consistency of the results, we construct the networks for different choices of (unweighted) cross-link density d_{sh} and internal link density $d_s = d_h = 2d_{sh}$ (adapted from Wiedermann et al. [2016a]).

hence, again separates well the dynamics within the troposphere from that inside the stratosphere (Fig. 6.8C). For all layers above 200 mbar $\mathcal{T}^*_{hs}$ becomes almost independent of the cross-link density ρ_{si} that is fixed when constructing the network. The same property also holds for the n.s.i. global cross-clustering coefficient $\mathcal{C}^*_{hs}$ computed over all nodes in the different HGT fields (Fig. 6.8D).

In agreement with the local measures discussed in Sec. 6.5 it is observed that n.s.i. cross-transitivity and n.s.i. global cross-clustering coefficients are in most cases larger for nodes in the HGT fields than for nodes in the SST field (compare Fig. 6.8A with Fig. 6.8C and Fig. 6.8B with Fig. 6.8D). As for the n.s.i. local cross-clustering

coefficients this indicates again the hierarchical network structure, i.e., a tendency for nodes in the HGT field to form triangular structures with nodes in the SST field, that is present across all atmospheric layers ranging from the troposphere to the lower stratosphere. The detailed structure of this hierarchy, however, seems to vary with the different atmospheric layers under study.

This observation further holds not only for the case of $d_{sh} = 0.005$ that was used in the previous sections but also for larger values that are chosen from a reasonable range (Fig. 6.8). Thus, also with respect to the assessment of global interacting network characteristics the results are considered to be sufficiently robust with respect to the choice of the networks' link densities.

In general, the quantitative and qualitative properties of the n.s.i. cross-transitivity and n.s.i. global cross-clustering coefficients vary with the different atmospheric layers. Hence, as for the single layer case (Chap. 4) these global characteristics again serve to inter-compare and distinguish between different correlation structures in a coupled climate network.

6.6. Summary

This chapter presented a detailed analysis of the correlation structure between atmospheric and ocean dynamics in the Northern Hemisphere extratropics from the viewpoint of coupled climate networks. A comparison between the n.s.i. cross-degree density (measuring the weighted share of significant correlations between grid points in different layers) and the leading mode of the maximum covariance analysis (MCA) revealed an expected high congruence between both methods for the considered data sets. However, coupled network analysis, and particularly the investigation of higher-order network parameters, allows to further disentangle the underlying correlation structure. Compared to MCA, the (average) n.s.i. cross-degree density in combination with the (average) n.s.i. local cross-clustering coefficient provided additional insights on areas in the ocean and the atmosphere that show strong mutual correlations as well as localized versus delocalized correlation structures with the respective opposite field. Specifically, it has been shown that nodes in the SST field tend to correlate with multiple mutually unconnected groups of similar nodes within the respective HGT fields and from investigating the interdependency between n.s.i. cross-degree density and n.s.i. local cross-clustering coefficient, it was found that the correlations between the ocean and the atmosphere exhibits a hierarchical structure in the sense of a power-law relationship between both properties. A visual inspection of the coupled climate network for the case of the 500 mbar HGT field revealed that the observed structure could be related with a forcing of the ocean by the dominant atmospheric patterns above the Atlantic and the Pacific. Ultimately, global network characteristics further supported the results obtained from their local cor-

respondents by showing that the observed structure is valid for large parts of the atmosphere ranging from the troposphere to the lower stratosphere.

In order to discriminate between the internal variability of the fields under study and possible influences of an external forcing, future work should analyze ensemble simulations of general circulation models to rule out common driver effects or assess the likelihood of their influence on the observed structures. In order to investigate the influence of spatio-temporal auto-correlation on the outcome of the present analysis the network could be alternatively constructed by estimating pairwise thresholds from surrogate data as proposed by Paluš et al. [2011]. This approach would, however, break the comparability of the network approach with that of maximum covariance analysis, such that a different way of validating and comparing the results must be found.

It remains to be noted, that the present analysis currently does not account for any possible effects induced by a common external forcing of the fields, which might artificially induce correlations and, hence, cause the presence of spurious links between nodes or triples of nodes. Also, indirect (partial) correlations or common driver effects within each of the fields have not yet been accounted for when constructing the coupled climate networks. Conditioning out these possible influences by means of information theoretic approaches [Runge et al., 2012a, 2013] and causal effect networks [Kretschmer et al., 2016, Runge et al., 2015] thus remains as a subject of future research.

To this end, the analysis presented here has only been performed for the pairwise correlation between one atmospheric layer and the ocean. Future studies should further explore the possibility to refine the interacting networks framework to also quantify interactions in a climate network consisting of more than two subnetworks. Specifically, when studying coupled climate networks in the Northern Hemisphere, one should also consider Arctic sea ice as an additional observable in the network construction. Its dynamics have already been discovered as an influencing factor on atmospheric teleconnections and land snow cover in the Northern Hemisphere [Handorf et al., 2015]. The study of coupled climate networks can help here to further disentangle and quantify possible changes in correlations between ocean and atmosphere over the course of the past decades that may have been induced by processes related to the Arctic amplification [Serreze and Francis, 2006]. Moreover, it is of great interest to apply methods from network theory not only to coupled networks composed of different climatic fields (as presented in this work), but also to networks constructed from just one single climatic field that divides into dynamically distinct areas [Hlinka et al., 2014] or communities [Steinhaeuser et al., 2011, Tsonis et al., 2010]. The framework presented in this work could then be utilized to study and quantify correlations between these detected or defined regions on or parallel to the Earth's surface. This would allow for a detailed investigation of correlation struc-

tures between different climatic subsystems such as, for example, the Indian Summer Monsoon and the El Niño Southern Oscillation as discussed in Chap. 4 and 5

This chapter concludes the analysis of functional climate networks that form the first part of this work. It has been shown in both applications that climate networks complement methods from classical statistical climatology as they unravel structures in the underlying similarity assessment that were previously undiscovered. In particular, global network characteristics based on the counting of triangular linkages between nodes, i.e., strongly connected clusters within the network have been highlighted as a powerful discriminator between different periods in the temporal evolution of a climate network (Chap. 4) or between subnetworks representing distinct layers in the atmosphere with respect to their interaction structure with the ocean (Chap. 6).

The following second part of this work continues to develop methods for discriminating sets of networks into different classes. First, Chap. 7 transfers the assessment of the special case of spatially embedded climate networks (as discussed up to now) to the general case of any spatially embedded, mostly real-world, network. Particularly, a general scheme is developed to disentangle the effects of a networks' spatial embedding from the assessment of macroscopic network characteristics, an issue that has in the previous chapters been dealt with by the assignment of area-specific node weights. From there, Chap. 8 ultimately moves on to a general quantitative assessment of the *complexity* of networks and derives a statistical measure which proves useful in the objective construction of functional (climate) networks.

Part II.

Methodological extensions & real-world networks

The first part of this work focused on the study of so-called functional climate networks. Specifically, it has been shown that network theory provides powerful tools to discriminate different phases in the evolution of the climate system, e.g., in terms of the El Niño Southern Oscillation (Chap. 4). Additionally, the study of coupled climate networks reveals differences in the structure of ocean-atmosphere interactions with respect to each of these two climatic subsystems (Chap. 6).

In addition to functional networks, where nodes represent entities of a larger dynamical system in either space and/or time and links denote some functional or statistical dependency between them, the second frequently studied class of networks is that of *structural* real-world networks [Bullmore and Sporns, 2009, Newman, 2003, Sporns et al., 2005]. Here, nodes directly represent physical entities such as intersections [Chan et al., 2011, Gastner and Newman, 2006], airports [Bagler, 2008] or individuals [Croft et al., 2008] and links mostly indicate physical or at least directly measurable relations between them. In the above examples, such relations could be roads, scheduled flights or friendship. This second part of the present work is devoted to the study of a variety of such systems.

Chapter 7 presents a framework to systematically analyze one special class of networks, namely spatially embedded networks, where nodes are assigned a specific location in some metric space. Usually, studies of such networks do not account for possible heterogeneities in the spatial distribution of nodes which may induce biases in the assessment of associated global and local network characteristics. While for climate networks such heterogeneities were easily addressed by the assignment of node-weights according to the latitudinal position of corresponding grid points (see, e.g., Chap. 3.1) a similar association of geographical location to node-weights is not as easily achieved for most real-world networks as their nodes are usually much more heterogeneously distributed in space. Alternatively, nodes may be weighted according to the area on the Earth's surface that they represent by means of a Voronoi-Tessalation of the underlying domain of interest [Rheinwalt et al., 2012]. However, this procedure only poorly accounts for boundary effects that arise if systems without periodic boundaries, such as smaller regions of the Earth, are studied. Also, in most cases it is more useful to assign nodes a weight according to other related quantities, such as the capacity of an airport or the GDP of a country [Heitzig et al., 2012, Maluck and Donner, 2015]. Hence, a node-weighting based on geographic location is often highly unfeasible. Addressing this issue, Chap. 7 introduces a hierarchy of null models to systematically study the effect of spatial embedding on a network's topological characteristics that are based on geographically constrained rewiring the structure of the network under study. The thus resulting surrogate networks can be compared with the original network which allows to quantify to what extent associated network characteristics are already predetermined by the spatial embedding of the nodes and links. Based on such an assessment a discrimination

in two classes is possible: (i) Networks whose macroscopic structure is to a large extent predetermined by the spatial embedding of the nodes alone and (ii) networks for which the corresponding macroscopic structure is largely independent from the spatial embedding.

Ultimately, this work closes by addressing a still open issue in complex network science. While being able to characterize the spatial and topological characteristics of networks by means of a variety of metrics it remains a task of ongoing research to quantify the actual *complexity* of such systems. Chap. 8 proposes a measure of statistical complexity that is inspired by approaches from nonlinear time series analysis and recurrence network analysis. It thus links back to the study of functional networks as presented in part I of this work. The statistical complexity is composed of a network's averaged per-node entropic measure and the associated Jenson-Shannon divergence. The study analyses 29 real-world networks and shows that networks of the same category tend to cluster in distinct areas of the resulting complexity-entropy plane. Again, linking back to the study of functional networks, it is finally shown that the proposed framework is useful to objectively construct threshold-based networks, such as functional climate networks or recurrence networks, by choosing the threshold such that the statistical network complexity is maximized.

Chapter 7.

Estimating effects of spatial embedding

Real-world networks with nodes embedded in a metric space have gained increasing interest in recent years [Barthélemy, 2011]. The effects of such a spatial embedding on the networks' structural characteristics, however, are rarely taken into account when studying their macroscopic properties. This chapter proposes a hierarchy of null models to generate random surrogates from a given spatially embedded network that can preserve global and local statistics associated with the nodes' embedding in a metric space. Comparing the original network's and the resulting surrogates' global characteristics allows to quantify to what extent these characteristics are already predetermined by the spatial embedding of the nodes and links. The resulting framework is applied to various real-world spatial networks and it is shown that the proposed models capture macroscopic properties of the networks under study much better than standard random network models that do not account for the nodes' spatial embedding. Depending on the actual performance of the proposed null models, the networks are ultimately categorized into different classes.

Section 7.1 provides a general introduction to the study of real-world spatial networks and highlights possible issues with the assessment of such systems which are addressed by this study. After some preliminary observations presented in Sec. 7.2, Sec. 7.3 introduces the algorithms behind the random network models proposed in this work. Section 7.4 gives an overview on the network data that is investigated and Sec. 7.5 showcases the results of the study. Section 7.6 again summarizes the results and provides an outlook on future directions of research.

This chapter is primarily based on the study that is presented in P5 [Wiedermann et al., 2016b].

7.1. Spatial networks

Many, if not most complex systems that exhibit a network structure are spatially embedded in some metric space [Barthélemy, 2011]. Examples of large interest include social [Capocci et al., 2006], transportation and distribution [Banavar et al., 1999, Chan et al., 2011, Jarvis et al., 2015], communication [Buldyrev et al., 2010], or electricity networks [Buzna et al., 2009]. As discussed before, one generally observed aim is to to classify such investigated networks into different categories, such as small-world networks [Watts and Strogatz, 1998] and subclasses thereof [Amaral et al., 2000, Barabasi, 2009]. Most studies, however, do not take into account the possible influence of a network's spatial structure on its resulting micro- or macroscopic characteristics [Bialonski et al., 2010]. Thus, it often remains unquantified whether a certain categorization of a network, such as the small-world property, is to some extent already explicable as emerging from the network's spatial embedding alone. Specifically, established random network models that may be used to assess whether a network follows a certain rule of construction solely preserve topological characteristics, such as the link density in *Erdős-Rényi random graphs* [Erdös and Rényi, 1960] or the degree sequence in the *configuration model* [Molloy and Reed, 1995].

To classify possible types of spatially embedded networks, several models have been proposed that generate random surrogates from a given spatially embedded node sequence by, e.g., randomly distributing links according to the spatial distance between nodes [Barnett et al., 2007, Heitzig et al., 2012], setting a prescribed linkage probability between nodes to address boundary effects in climate networks [Rheinwalt et al., 2012] or optimizing the length-dependent costs related to the construction of a link in power grids [Schultz et al., 2014]. A variety of further studies introduced random network models to specifically reproduce statistics associated with brain networks. Most of these models are of generative nature and set up artificial networks that are then compared with observations made from data. Some models are designed such that connections between regions of similar input are favored but long-range connections are penalized [Vertes et al., 2012], e.g., by optimizing the interplay between range-dependent wiring cost and processing efficiency [Chen et al., 2013]. In other models the connectivity between nodes or areas of the brain depends on gravitational forces between these different areas [Song et al., 2014] and even other models are designed such that links are put between nodes depending on their participation coefficient [Rubinov et al., 2015].

All of the above mentioned models, however, have been primarily designed to assess and reproduce construction principles behind certain types of complex networks and their underlying mechanism are usually tailored to a specific application.

In order to explicitly study the general influence of a network's spatial embedding on its resulting macroscopic characteristics, this chapter proposes a set of random

network models to create surrogates that preserve certain geographical and topological features of these given networks. The surrogates are constructed by iteratively rewiring the original network while preserving a set of its geographical features. In particular, the first model termed *GeoModel I* preserves the global link-length distribution in addition to the degree sequence. A second model referred to as *GeoModel II* additionally preserves for each node the length distribution of the links connected to it and, hence, imposes an even stronger spatial constraint on the rewiring process. The resulting surrogate networks allow for evaluating to what extent observed macroscopic properties of a given network are explicable by geometric constraints inflicted on the network's structure, while no assumptions on the specific construction principles are necessary.

After a thorough derivation, the method is applied to a number of real-world complex networks: the US airline network, the US interstate network, the Internet [Gastner and Newman, 2006], the Scandinavian power grid [Menck et al., 2014], a world trade network [Lenzen et al., 2012], and the road network of a German city. Additionally, the proposed models are applied to a random geometric graph [Penrose, 2003] and an Erdős-Rényi network [Erdös and Rényi, 1960]. For comparison, iteratively rewired surrogate networks that only preserve topological characteristics of the given networks, namely the mean degree on the one hand and the degree distribution on the other hand, are constructed as well.

To characterize the macroscopic structure of the networks under study as well as their corresponding random surrogates, this work relies on two commonly used global network measures, the global clustering coefficient $\mathcal{C}$ (Eq. (2.9)) and the average path length $\mathcal{L}$ (Eq. (2.10)) [Newman, 2003]. Recall from Sec. 2.3 that an evaluation of both measures is commonly used to classify a network under study as, e.g., a small-world network, which is defined to display a high clustering coefficient while at the same time showing a low average path length [Davidsen et al., 2002, Watts and Strogatz, 1998].

Obviously, these two metrics form only a small subset of possible measures that could be investigated in the course of this work. Among others, link-weighted quantities such as the weighted average shortest path length [Latora and Marchiori, 2003] or the weighted clustering coefficient [Newman, 2001] have become of increasing interest when quantifying the topology of real-world complex networks [Barrat et al., 2004a]. However, as this work aims to inter-compare resulting quantities among different networks under study, one would have to define a concise weighting scheme that is applicable simultaneously to all of them. This usually raises conceptual issues as on the one hand in airline networks links are commonly weighted by the number of available seats or the number of flights on a route between two airports [Bagler, 2008, Barrat et al., 2004b]. On the other hand road networks may be weighted by, e.g., the total traffic flow along certain roads [De Montis et al., 2007] while trade net-

works are weighted by the total trade volume between two partners [Fagiolo et al., 2008]. Thus, macroscopic quantities computed from different networks weighted according to these definitions would not be straightforward to compare as they describe different physical entities. Even neglecting these usually employed definitions of link weights, a consistent weighting that is purely based on distances is similarly hard to perform. While infrastructure networks (such as airline networks) should be weighted linearly according to the spatial distance between nodes, link weights in communication networks (such as the Internet) should correspond to the inverse distance, i.e., the per-link efficiency [Latora and Marchiori, 2003]. For networks such as the world trade network or synthetic networks as the Erdős-Rényi graph and the random geometric graph a weighting based on node-distances would further lack a proper and concise interpretation.

Thus, this work is restricted to the assessment of well established unweighted network measures which were already successfully applied to quantify and classify the topology of spatially embedded networks [Crucitti et al., 2006, Gastner and Newman, 2006]. This procedure provides results that are easily comparable and interpretable among the different networks studied in this work.

7.2. Preliminaries

Each node in the network $G = (V, E)$ under study is assigned a position $\mathbf{x}_i$ in some metric space of dimension ν. In the applications presented in this work, nodes are either embedded on the surface of a sphere, i.e., the Earth's surface, or in a Cartesian coordinate system. In the first case, the position of a node is determined by its latitudinal and longitudinal coordinates λ_i and ϕ_i and, hence, $\mathbf{x}_i = (\lambda_i, \phi_i)$. In the second case, a node's position is given by its Cartesian coordinates x_i and y_i with respect to some arbitrarily chosen origin, $\mathbf{x}_i = (x_i, y_i)$. The $N_x \times N_x$ distance matrix $\mathbf{D}$ then gives the distance between all nodes in the network. For the case of a spherical coordinate system, its entries D_{ij} are computed as the great circle distances between nodes,

$$D_{ij} = r \arccos(\sin\lambda_i \sin\lambda_j + \cos\lambda_i \cos\lambda_j \cos(\phi_j - \phi_i)). \tag{7.1}$$

r denotes the radius of the sphere, which is rescaled to unit length in all applications, $r = 1$. For the case of a Cartesian coordinate system, the entries of $\mathbf{D}$ are given by the Euclidean distance between two nodes,

$$D_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}. \tag{7.2}$$

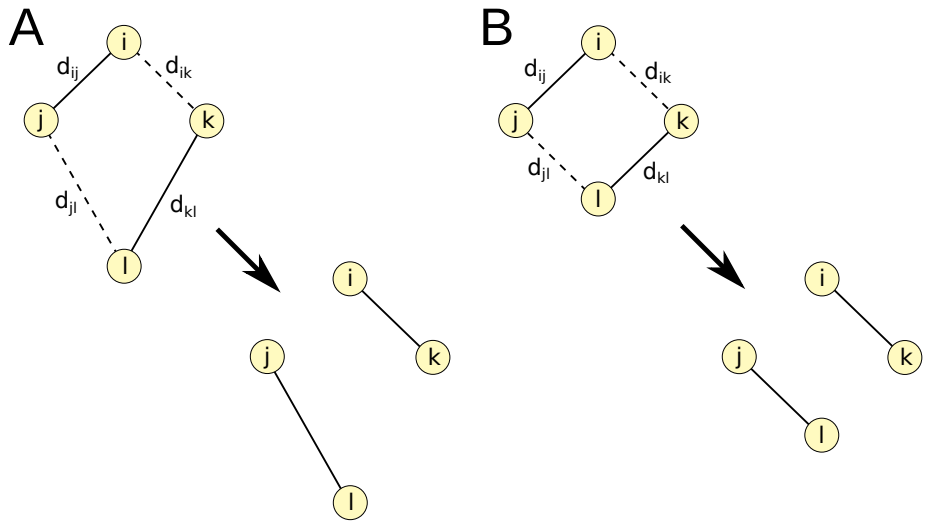

Figure 7.1.: Sketch of the rewiring process that generates randomized surrogates of a given original network by applying GeoModel I (A) and GeoModel II (B). Nodes i, j, k and l are drawn at random, such that i is linked to j and k is linked to l (solid lines) but no links are present between i and k, and j and l (dashed lines). According to the chosen network model the distances $D_{\bullet\bullet}$ between the nodes are evaluated. If the nodes form an approximate kite (A) or diamond (B) with the connections between them present as depicted, previous links are replaced by links connecting i with k and j with l (adapted from Wiedermann et al. [2016b]).

From this the local cumulative distribution function $F_i(l)$ of link lengths l of node i follows directly as

$$F_i(l) = \frac{1}{k_i} \sum_j A_{ij} \Theta(l - D_{ij}). \tag{7.3}$$

Here, k_i denotes the degree of node i (Sec. 2.2) and $\Theta(\cdot)$ is the Heaviside function. The global cumulative link length distribution $F(l)$ in the network is then given as

$$F(l) = \frac{1}{2M} \sum_i k_i F_i(l), \tag{7.4}$$

with $M = |E|$ denoting the total number of links in the network.

7.3. Random network models

The random network surrogates are constructed from a given network by applying four different algorithms. Two of them, random link switching and random rewiring do not take into account any spatial embedding of the network's nodes, whereas this consideration is an explicit part of the two novel models, GeoModel I and GeoModel

II.

The general structure of the algorithms is described as follows: Starting from a copy $\mathbf{A}'$ of the original network's adjacency matrix $\mathbf{A}$:

(i) Draw four distinct nodes i, j, k, l uniformly at random from V.

(ii) Depending on the applied random network model under study, check whether a certain condition $\mathcal{E}$ is TRUE. If $\mathcal{E}$ is FALSE return to step (i).

(iii a) (applies to random rewiring) If $\mathcal{E}$ is TRUE, break the link connecting i with j and establish a new link connecting k with l. Hence, $A'_{ij} = A'_{ji} = 1 \to 0$ and $A'_{kl} = A'_{lk} = 0 \to 1$.

(iii b) (applies to all other random network models) If $\mathcal{E}$ is TRUE, break the links connecting i with j and k with l and establish links connecting i with k and j with l. Hence, $A'_{ij} = A'_{ji} = 1 \to 0$, $A'_{kl} = A'_{lk} = 1 \to 0$, $A'_{ik} = A'_{ki} = 0 \to 1$, and $A'_{jl} = A'_{lj} = 0 \to 1$.

(iv) As long as a certain number of rewirings m is not reached return to (i) with the modified adjacency matrix $\mathbf{A}'$.

The resulting modified copy $\mathbf{A}'$ of the original network's adjacency matrix $\mathbf{A}$ is then returned for further evaluation. The following sections introduce the explicit forms of the conditions $\mathcal{E}$ for a rewiring process to take place.

Random rewiring

Random rewiring, the simplest case, takes place if a link exists between the randomly drawn nodes i and j, but no link exists between k and l. Hence,

$$\mathcal{E} = A'_{ij} \wedge \neg A'_{kl}. \tag{7.5}$$

Here, $\wedge$ denotes the truth-functional operator of logical conjunction and $\neg$ denotes the logical complement, i.e., $\neg 1 = 0$ and $\neg 0 = 1$. Note that in the scope of this work the adjacency matrices' entries $A_{ij} = 1$ are interpreted as the logical 1 (TRUE) and entries $A_{ij} = 0$ are interpreted as the logical 0 (FALSE). Thus, the resulting value of $\mathcal{E}$ is also a logical value being either TRUE or FALSE.

The definition of $\mathcal{E}$ is then plugged into step (ii) and (iii a) in the above algorithm and depending on its value a new set of four nodes is drawn or a rewiring process takes place. Random rewiring solely preserves the average degree $K = N_x^{-1} \sum_i k_i$ in the network and, hence, after sufficiently many rewiring steps, the resulting surrogate network converges to an Erdős-Rényi random graph [Erdös and Rényi, 1960].

Random link switching

In addition to the mean degree K, random link switching preserves the local degree of each node in the network as well, but still neglects any aspect of a network's spatial embedding [Artzy-Randrup and Stone, 2005, Zamora-López et al., 2008]. This framework has in the past already been successfully applied to study topological properties of real-world systems such as the Internet [Maslov et al., 2004] or protein networks [Maslov and Sneppen, 2002]. Here, for the four nodes drawn in step (i) of the construction algorithm, one needs to ensure that i is linked with j and k is linked with l, but i and k as well as j and l are not yet connected (Fig. 7.1). Hence, the condition $\mathcal{E}$ reads

$$\mathcal{E} = \mathcal{E}_1 = A'_{ij} \wedge A'_{kl} \wedge \neg A'_{ik} \wedge \neg A'_{jl}. \tag{7.6}$$

As the degree of each node is preserved, the resulting surrogate networks relate to the results one would obtain from the configuration model [Molloy and Reed, 1995]. However, for the present case the surrogate networks display no self-loops or multiple links between nodes.

GeoModel I

In addition to the above criterion $\mathcal{E}_1$, GeoModel I aims to also preserve the global link length distribution $F(l)$. Hence, the potentially newly established links must be of the same length as those that are removed from the network. This means that the four randomly drawn nodes i, j, k and l must form a kite with exactly one link present at each of the two sides of the same length (Fig. 7.1A). Since the nodes are usually embedded in a continuous domain this equality can only be fulfilled up to a certain accuracy. Hence, it is demanded that the newly established links have approximately the same length as the existing ones with some tolerance ϵ. In other words, the difference in lengths between the present and potentially established links should not exceed a certain fraction ϵ of the existing links' lengths. Thus, the following condition must be fulfilled,

$$\mathcal{E} = \mathcal{E}_1 \wedge \mathcal{E}_2 \text{ with} \tag{7.7}$$

$$\mathcal{E}_2 = \Theta(\epsilon d_{ij} - |d_{ij} - d_{ik}|) \wedge \Theta(\epsilon d_{kl} - |d_{kl} - d_{jl}|). \tag{7.8}$$

Here, ϵ measures the maximum allowed relative deviation in length between the existent and potentially newly established links. As for the entries of the adjacency matrices $\mathbf{A}$, a value of $\Theta(\cdot) = 1$ here denotes the logical 1 (TRUE) and $\Theta(\cdot) = 0$ correspondingly represents the logical 0 (FALSE).

GeoModel I preserves the degree distribution in the same way as random link switching, but in addition approximately preserves the global link length distribution

$F(l)$. The ensemble Ω_{GMI} of possible surrogates constructed by GeoModel I therefore forms a subset of the ensemble Ω_{rls} of all those surrogates possibly constructed from random link switching, $\Omega_{GMI} \subseteq \Omega_{rls}$. Generally, it is to be expected that with an increasing ϵ the Hamming distance $\mathcal{H}$ (Eq. (2.15)) between the original networks and its surrogates increases. However, an increase in ϵ also induces larger deviations between the link length distributions of the original and surrogate networks. Hence, the maximum meaningful value of ϵ is estimated by using a Kolmogorov-Smirnoff (KS) test [Press et al., 1996]. This test is based on the original network's and the surrogate network's distribution of link lengths, $F(l)$ and $F'(l)$, respectively. The corresponding KS-statistic $\mathcal{K}$ is given as

$$\mathcal{K} = \max_{0<l<\infty} |F(l) - F'(l)|. \tag{7.9}$$

The two distributions are considered statistically indistinguishable at a confidence level α if [Press et al., 1996]

$$Q_{KS}\left(\left[\sqrt{M}_e + 0.12 + 0.11/\sqrt{M}_e\right]\mathcal{K}\right) > \alpha. \tag{7.10}$$

Here, $M_e = MM'/(M + M') = M/2$ is the effective number of links constituting each distribution and Q_{KS} is given as

$$Q_{KS}(x) = 2\sum_{j=1}^{m}(-1)^{j-1}\exp(-2j^2x^2). \tag{7.11}$$

In theory, the above sum has infinitely many entries, $m = \infty$. In this work it is $m = 100$ to obtain an acceptable approximation. In order to estimate an optimal value of ϵ, it is demanded that for an ensemble of N_{sur} surrogate networks the resulting link length distributions are statistically indistinguishable from that of the original network in 95% of all cases under a confidence level of $\alpha = 95\%$. In other words, Eq. (7.10) must hold in 95% of all cases for $\alpha = 95\%$.

GeoModel II

In order to not only preserve the global but also the local (per-node) link length distributions $F_i(l)$, it must be demanded that the two links to be removed and the two links to be established all have approximately the same length. Hence, the nodes i, j, k and l form a diamond. That way, none of the lengths of links emerging from either of the four nodes is changed under each rewiring step. As above, in most situations this criterion can only be fulfilled approximately and for convenience, the

Name	N_x	M	K	d	$\mathcal{C}$	$\mathcal{L}$	ϵ_I	ϵ_{II}	Grid
US airline	190	837	8.86	0.0466	0.679	2.176	0.04	0.07	S
Internet	13.372	28.253	4.23	0.0003	0.423	3.630	0.04	0.01	S
US interstate	935	1.315	2.82	0.0030	0.107	20.207	0.17	0.24	S
Scandinavian power grid	236	318	2.71	0.0115	0.084	9.156	0.16	0.27	C
World trade	186	7.043	76.14	0.4094	0.815	1.594	0.02	0.04	S
Urban roads (Eschwege)	855	1.174	2.75	0.0032	0.050	18.313	0.15	0.22	S
Random geometric graph	2.000	5.493	5.50	0.0027	0.588	30.428	0.11	0.13	C
Erdős-Rényi graph	2.000	5.493	5.50	0.0027	0.003	4.643	0.01	0.01	C

Table 7.1.: Overview of all networks investigated in this chapter including their number of nodes N_x and links M, average degree K, link density ρ, global clustering coefficient $\mathcal{C}$, and average path length $\mathcal{L}$. ϵ_I and ϵ_{II} denote the relative tolerances that are chosen for generating random network surrogates from GeoModel I and GeoModel II, respectively. The last column indicates the type of the underlying grid, which is either represented in spherical (S) or cartesian (C) coordinates.

same parameter ϵ as above is utilized to extend the previous conditions $\mathcal{E}_1$ and $\mathcal{E}_2$ as

$$\mathcal{E} = \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge \mathcal{E}_3 \text{ with} \tag{7.12}$$

$$\mathcal{E}_3 = \Theta(\epsilon \max(d_{ik}, d_{jl}) - |d_{ik} - d_{jl}|). \tag{7.13}$$

Thus, the difference in length of the newly established links (and therefore also the difference in lengths of the existing links) must not be larger than a certain fraction ϵ of their respective maximum length. For this study the maximum of D_{ik} and D_{jl} is taken as the reference scale of the tolerance window. However, other choices, such as the minimum value or the arithmetic mean of the two, might also be possible and would result in different optimal values of the tolerance parameter ϵ. A detailed investigation on the effect of the actual definition of the link length that is chosen as a reference, however, is suggested as a subject of future research.

Again, the ensemble Ω_{GMII} of all possible surrogates constructed from GeoModel II forms a subset of all possible surrogates constructed from GeoModel I and random link switching, since it only imposes a further condition in addition to the already employed ones. Hence, $\Omega_{GMII} \subseteq \Omega_{GMI} \subseteq \Omega_{rls}$.

7.4. Data

This study considers different real-world networks to illustrate the performance of the above algorithms and test to what extent macroscopic characteristics are recaptured by random network surrogates that take into account spatial constraints on the distribution of links in the network. First three different previously studied infrastructure networks are investigated [Gastner and Newman, 2006]: the US

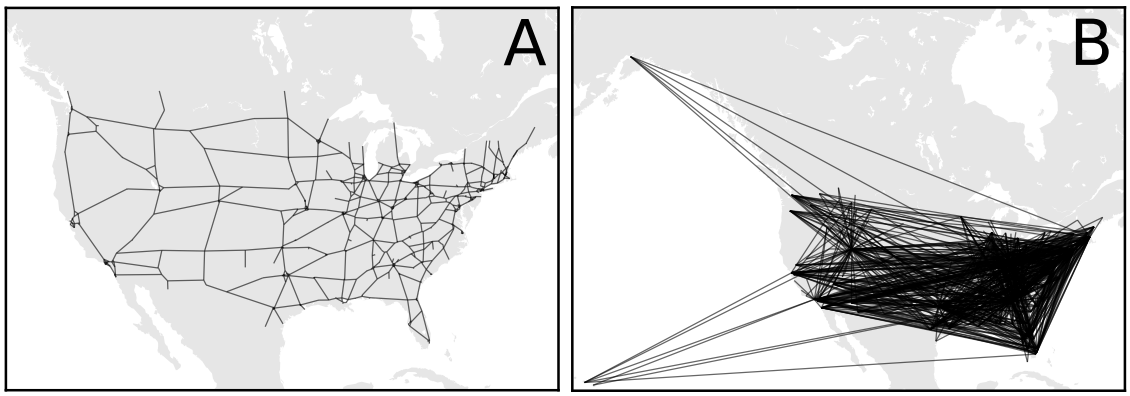

Figure 7.2.: Selection of two networks that are investigated in this study: (A) The US interstate and (B) the US airline network.

airline network with nodes displaying airports and links indicating flights scheduled between them, the US interstate network with links representing highways and nodes serving as country borders, termination points and intersections between highways and the physical layer of the Internet where links stand for data connections between autonomous systems around the globe, i.e., the corresponding nodes. Contrasting the case of the interstate network, this work also studies an infrastructure network of smaller spatial scale by retrieving the urban road network of a small Germany town (Eschwege) from `www.openstreetmap.org` (accessed 2012-01-30). Here, nodes again represent intersections and links are roads connecting them. Moreover, the framework is applied to the Scandinavian power grid, where links represent high voltage transmission lines and nodes are substations representing net consumers and producers [Menck et al., 2014]. The latter type of networks has been intensively studied in the framework of complex network theory and the understanding of its global properties has been reported as crucial since these strongly determine their local behavior, e.g., the robustness to failures of single nodes [Albert et al., 2004, Crucitti et al., 2004, Kinney et al., 2005]. Finally, this study includes the world trade network of 2009 with nodes representing countries and links indicating trade between them [Lenzen et al., 2012] as a representative of a non-physical, yet spatially embedded transaction network. For later reference, Fig. 7.2 shows as an example the US interstate and the US airline network.

For comparison with these real-world networks, this study also investigates synthetic networks with known properties, which serve as a benchmark for the analysis. Particularly, a random geometric graph with nodes put randomly on a plain unit square is considered [Donges et al., 2012, Herrmann et al., 2003]. Here, all nodes with a spatial distance of less than 0.03 are connected to yield a manageable density of links. This network's macroscopic properties are expected to only be explainable

by considering random network models that take into account the spatial embedding of the nodes. For the sake of comparison, this work additionally investigates one realization of an Erdős-Rényi random graph with the same number and position of nodes and the same number of links randomly put between them as in the random geometric graph. As links are put without any relation to spatial distances, the simplest network model, i.e., random rewiring, should already capture this network's macroscopic features.

A summary of all networks included in this study together with each network's number of nodes N_x and links M as well as further network parameters is presented in Tab. 7.1.

7.5. Results

The four random network models introduced above are now applied to the different real-world and synthetic networks under study. A first step illustrates how to estimate a proper tolerance parameter ϵ for GeoModel I and GeoModel II. Specifically, the procedure is shown for the example of the US interstate network and the application of GeoModel I. Afterwards, the results of all four network models applied to the US interstate and the airline network are discussed in detail and it is shown to what extent macroscopic network characteristics are reproduced by the different network models. Finally, a comprehensive intercomparison between all networks investigated in this study by applying the different models to each real-world network is presented. It is evaluated, which macroscopic features of a network can be reproduced by which model and the real-world networks are assigned to different classes, i.e., those for which spatial embedding plays a minor role when estimating macroscopic properties and those where the spatial structure explicitly needs to be taken into account.

For all cases discussed from now on, an ensemble of $N_{sur} = 100$ surrogate networks is constructed for each network under study and each random model is iteratively rewired for $m = 20M$ steps.

Selection of the tolerance parameter ϵ

The only free parameter in GeoModel I and GeoModel II is the tolerance parameter ϵ in Eqs. (7.8) and (7.13), which determines which link lengths are treated as being sufficiently similar. To illustrate the influence of ϵ on the results presented in this chapter, GeoModel I is applied to the US interstate network (as shown in Fig. 7.2A) and $N_{sur} = 100$ surrogate networks are created so that they display the same degree sequence and approximately the same global link length distribution $F(l)$ as the original network. Figure 7.3 shows the average evolution of global clus-

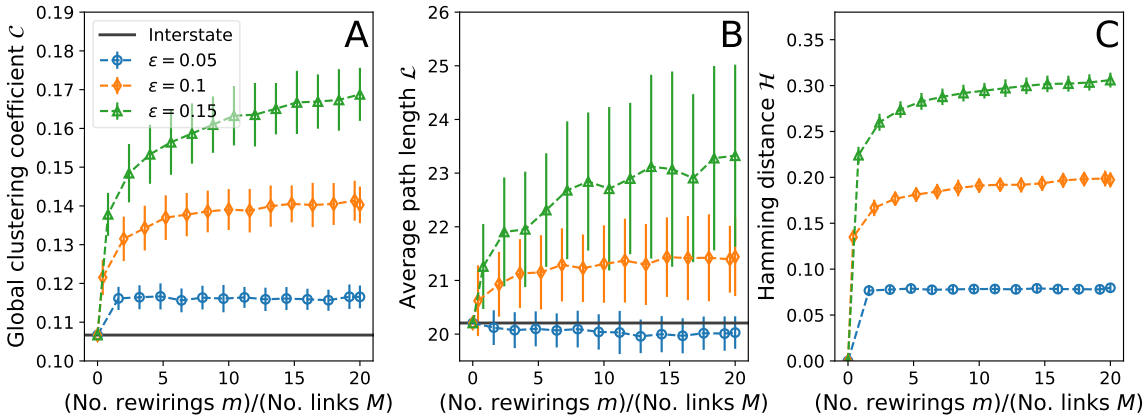

Figure 7.3.: Evolution of the global clustering coefficient $\mathcal{C}$ (A) and average path length $\mathcal{L}$ (B) with the number of rewirings for an ensemble of $N_{sur} = 100$ surrogates generated from the interstate network by applying GeoModel I and using different tolerances ϵ_I. Solid lines indicate the respective value of $\mathcal{C}$ and $\mathcal{L}$ of the interstate network itself. (C) Hamming distance $\mathcal{H}$ between the surrogate networks and the original network. Scatter symbols denote the mean value and error bars indicate one standard deviation of each measure (adapted from Wiedermann et al. [2016b]).

tering coefficient $\mathcal{C}$ (Eq. (2.9)), average path length $\mathcal{L}$ (Eq. (2.10)) and Hamming distance $\mathcal{H}$ (Eq. (2.15)) for different choices of ϵ in the corresponding surrogates. As expected, one notes that for the lowest choice of ϵ ($\epsilon = 0.05$) the surrogate networks' global clustering $\mathcal{C}$ and average path length $\mathcal{L}$ are closest to the values of the original network (Fig. 7.3A,B). However, in that case, the Hamming distance displays low values around 0.075, meaning that 92.5% of links in the original network are also present in the surrogate networks (Fig. 7.3C). With increasing ϵ the values of $\mathcal{H}$ also increase and, hence, the surrogate networks become increasingly dissimilar from the original network. At the same time, $\mathcal{C}$ and $\mathcal{L}$ also differ more from their target values (Fig. 7.3A,B).

As GeoModel I aims to approximately preserve the global link length distribution $F(l)$, the distribution of the KS statistics $\mathcal{K}$ for the ensemble of surrogate networks at different tolerance parameters ϵ is additionally examined (Fig. 7.4A). For low values of ϵ, all cumulative link length distributions are statistically indistinguishable with 95% confidence, which results in values of $\mathcal{K}$ being smaller than the critical value $\mathcal{K}_{crit}$. This value indicates the upper bound of the confidence interval (Fig. 7.4A) and is determined as the largest possible value that satisfies Eq. (7.10). However, as already discussed above, the Hamming distance $\mathcal{H}$ becomes very low for low ϵ and only a few links differ between the original and the surrogate networks (Fig. 7.4B). On the other hand, for large ϵ most link length distributions are dissimilar under the desired confidence level and, hence, the purpose of GeoModel I is not fulfilled. It is observed, that for $\epsilon = 0.17$, 95% of all distributions are statistically equivalent

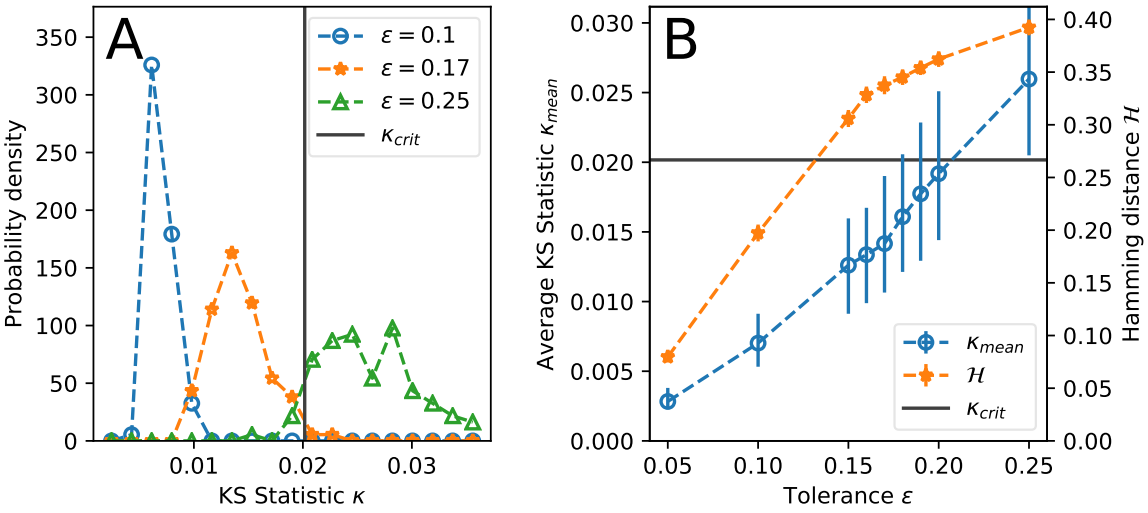

Figure 7.4.: (A) Distribution of KS statistics $\mathcal{K}$ measuring the maximum distance in the global cumulative link length distribution $F(l)$ between the interstate network and each of the $N_{sur} = 100$ network surrogates obtained by applying GeoModel I under different tolerances ϵ and $20M$ rewirings. (B) Average KS statistics $\mathcal{K}_{mean}$ and Hamming distance $\mathcal{H}$ after $20M$ successful rewirings depending on the choice of tolerance ϵ. Error bars denote one standard deviation for the Hamming distance and the 5th and 95th percentile of the distribution of KS statistics. The solid line indicates the critical value $\mathcal{K}_{crit}$ below which the surrogates' and the original network's link length distributions are considered statistically indistinguishable under a confidence level of $\alpha = 0.95$ (adapted from Wiedermann et al. [2016b]).

with 95% confidence and, hence, GeoModel I achieves its highest possible Hamming distance (Fig. 7.4).

Following the same procedure, the optimal tolerance can be obtained for GeoModel II as well as for all other networks under study. It is important to note that the values of ϵ generally differ between the two random network models as for GeoModel II the additional criterion $\mathcal{E}_3$ must be fulfilled. Therefore ϵ_I denotes the optimal tolerance for GeoModel I and ϵ_{II} the respective optimal tolerance for GeoModel II. A summary of all tolerances for each network and random network model is given in Tab. 7.1. Note that the obtained values of ϵ differ between $\epsilon = \epsilon_{II} = 0.01$ for the Internet and $\epsilon = \epsilon_{II} = 0.27$ for the power grid. Moreover, in most cases it holds that $\epsilon_I < \epsilon_{II}$.

Interstate network

With the two tolerances $\epsilon_I = 0.17$ and $\epsilon_{II} = 0.24$ estimated for applying GeoModel I and GeoModel II to the interstate network (Fig. 7.2A), this section now investigates the evolution of $\mathcal{C}$ and $\mathcal{L}$ with an increasing number of rewiring steps for the four different random network models (Fig. 7.5). Generally, the random rewiring and random link switching converge towards a state where there is hardly any further

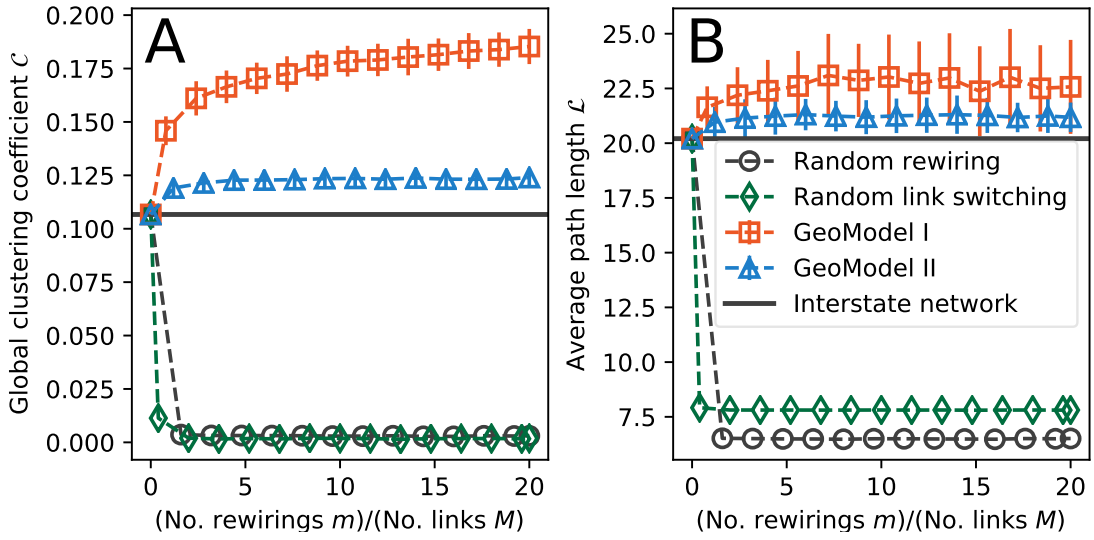

Figure 7.5.: Evolution of (A) global clustering coefficient $\mathcal{C}$ and (B) average path length $\mathcal{L}$ with the number of rewirings averaged over ensembles of $N_{sur} = 100$ surrogates generated from the interstate network by applying the different random network models (dashed lines). For GeoModel I and GeoModel II the tolerances are set to $\epsilon_I = 0.17$ and $\epsilon_{II} = 0.24$, respectively. Scatter symbols denote the mean value of each measure. Error bars indicate one standard deviation and are shown if their size exceeds that of the symbol. Solid lines indicate the value of $\mathcal{C}$ and $\mathcal{L}$ in the original network (adapted from Wiedermann et al. [2016b]).

fluctuation in the evolution of $\mathcal{C}$ and $\mathcal{L}$ after less than $2M$ steps of rewiring (Fig. 7.5). Similarly, GeoModel II converges after $5M$ steps. Only for GeoModel I, one notes small fluctuations in the average evolution of $\mathcal{L}$ (Fig. 7.5B) and a slow saturation in the average evolution of $\mathcal{C}$ (Fig. 7.5A) up to the maximum value of $m = 20M$ steps of rewiring.

Remarkably, surrogate networks obtained from random rewiring and random link switching do not capture well the macroscopic characteristics of the interstate network indicated by large deviations of $\mathcal{C}$ and $\mathcal{L}$ from their original values (Fig. 7.5A,B). In fact, with respect to the global clustering coefficient $\mathcal{C}$, the two models perform equally badly (Fig. 7.5A). For the average path length $\mathcal{L}$, the additional constraint of a preserved degree sequence when applying random link switching yields a slight improvement over the process of random rewiring as in average the surrogate networks' values of $\mathcal{L}$ are closer to that of the original network.

Additionally taking into account spatial constraints on the lengths of links in the random networks and, hence, applying GeoModel I and GeoModel II yields macroscopic characteristics of the surrogates that are much closer to those of the original network (Fig. 7.5). Specifically, GeoModel I already gives values of $\mathcal{L}$ very close to that of the original interstate network (Fig. 7.5B), while the resulting values of $\mathcal{C}$ still deviate strongly from their target (Fig. 7.5A). The additional constraint of

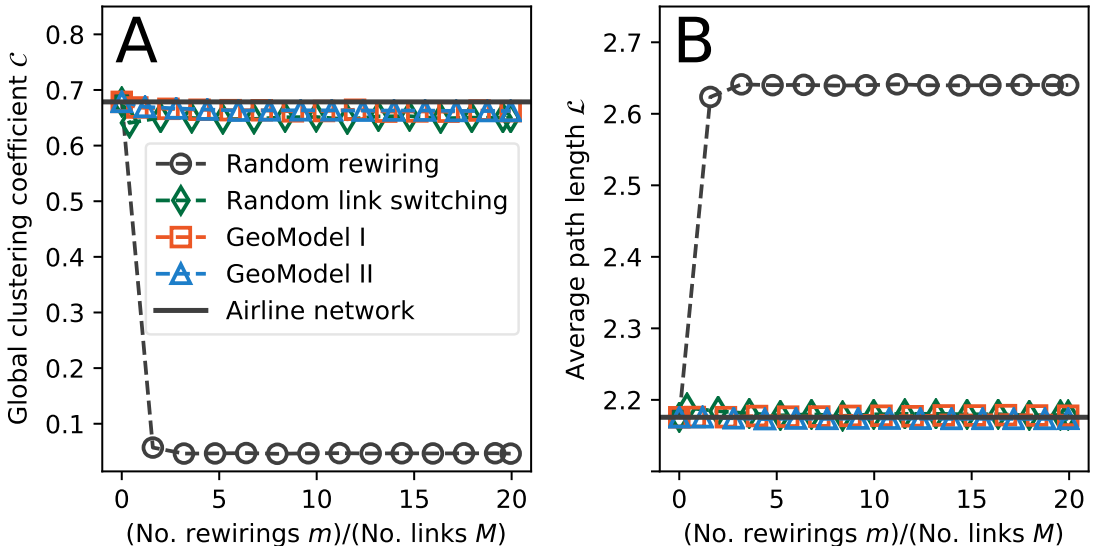

Figure 7.6.: Same as Fig. 7.5 but for the airline network and tolerances of $\epsilon_I = 0.04$ and $\epsilon_{II} = 0.06$. Error bars are not shown as they do not exceed the size of the symbols (adapted from Wiedermann et al. [2016b]).

a preserved local link length distribution $F_i(l)$ overcomes this issue and GeoModel II provides surrogate networks that, in addition to $\mathcal{L}$, also approximate $\mathcal{C}$ in good agreement with the original network. However, slight differences in the two quantities obtained from GeoModel II in comparison with the original network's characteristics are still present. Additionally constraining the algorithm to also preserve a network's degree correlation [Zamora-López et al., 2008] might further improve the agreement between the surrogates and the original network. An investigation of such higher order effects remains as a subject of future research. It is also noted that GeoModel I and GeoModel II tend to give too high values of $\mathcal{L}$ and $\mathcal{C}$ at least for the particular case of the interstate network. This effect could be related to optimization principles, such as the minimization of intersection crossings for road networks, underlying the original network that are not accounted for by the surrogate networks' construction mechanism. Future studies should address in more detail, for what types of networks GeoModel I and GeoModel II possibly show systematic positive or negative biases with respect to the target values of, e.g., global clustering coefficient and average path length.

Airline network

The same procedure as discussed before is now applied to the airline network (Fig. 7.2B) and the evolution of global clustering coefficient $\mathcal{C}$ and average path length $\mathcal{L}$ with an increasing number of rewiring steps are computed for the four different random network models (Fig. 7.6). One notes a fast convergence towards a state with hardly

any further fluctuations in the average evolution of $\mathcal{C}$ and $\mathcal{L}$ for all four random network models. As for the interstate network, random rewiring does not produce surrogate networks that capture well the macroscopic characteristics of the airline network. However, in contrast to the former case, random link switching already reproduces very well both macroscopic quantities $\mathcal{C}$ and $\mathcal{L}$ (Fig. 7.6). GeoModel I and GeoModel II slightly improve these results further, but for the present case of the airline network a prescribed degree sequence already produces surrogate networks with properties close to those of the original network. Thus, for the airline network the spatial embedding of the nodes and the resulting characteristic distribution of link lengths is less important for its macroscopic properties as compared to the US interstate network.

Inter-comparison between different spatial networks

As in the previous sections, the evolution of global clustering coefficient $\mathcal{C}$ and average path length $\mathcal{L}$ is now computed for each of the networks under study (Tab. 7.1) by evaluating ensembles of $N_{sur} = 100$ realizations of each network model and using $m = 20M$ rewiring steps. To give a comprehensive summary, the average relative deviation $\Delta\mathcal{C}$ and $\Delta\mathcal{L}$ from the respective original network's values is computed for each network and network model,

$$\Delta\mathcal{C} = \frac{\langle \mathcal{C}_{sur} \rangle - \mathcal{C}_{orig}}{\mathcal{C}_{orig}} \tag{7.14}$$

$$\Delta\mathcal{L} = \frac{\langle \mathcal{L}_{sur} \rangle - \mathcal{L}_{orig}}{\mathcal{L}_{orig}}. \tag{7.15}$$

Figure 7.7 summarizes the results for all spatial networks under study. As expected, the Erdős-Rényi network's topological features are already well reproduced by random rewiring, while this is not the case for all other networks (Fig. 7.7A). In all cases, the global clustering coefficients $\mathcal{C}$ of the surrogate networks are lower than those of the respective original networks (indicated by negative values of $\Delta\mathcal{C}$ in Fig. 7.7A), which is in accordance with the fact that networks generated from random rewiring are expected to display a clustering coefficient close to their link density [Albert and Barabási, 2002]. Remarkably, the average path length of the world trade network is also already well reproduced by random rewiring (resulting in $\Delta\mathcal{L}$ close to zero in Fig. 7.7A), which is likely due to its large link density of $d \approx 0.4$. Note that for the Internet and the airline network, the randomly rewired surrogates have a positive bias for the average path length $\mathcal{L}$, while for all remaining networks, a negative bias is found.

As discussed in Section 7.5, the process of random link switching reproduces well the macroscopic properties of the airline network (Fig. 7.7B). The same observation

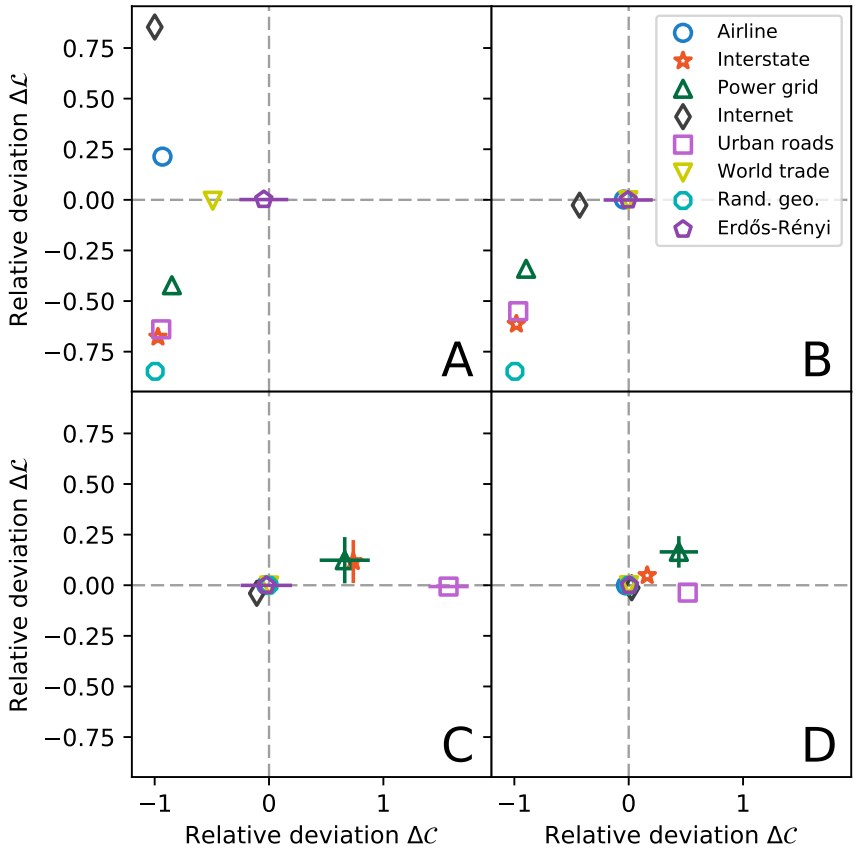

Figure 7.7.: Average relative deviation of global clustering coefficient $\Delta\mathcal{C}$ and average path length $\Delta\mathcal{L}$ from the respective original values computed over an ensemble of $N_{sur} = 100$ surrogate networks after $20M$ successful rewirings by applying (A) random rewiring, (B) random link switching, (C) GeoModel I and (D) GeoModel II. The tolerances ϵ_I and ϵ_{II} used for each network and random network model are given in Tab. 7.1. Error bars denote the standard deviation in $\Delta\mathcal{C}$ and $\Delta\mathcal{L}$ and are shown if their size exceeds that of the corresponding symbol (from Wiedermann et al. [2016b]).

also holds for the world trade and, as expected, for the Erdős-Rényi network. Additionally, the average path length of the Internet is already well captured by random link switching, too. Thus, the topological features of these networks are already well expressed in terms of their degree distribution and the spatial embedding of the nodes seems to have little influence on the average path length $\mathcal{L}$ and the global clustering coefficient $\mathcal{C}$. For the four other networks (power grid, urban roads, interstate and random geometric graph), only slight improvements are visible when comparing the relative deviations $\Delta\mathcal{C}$ and $\Delta\mathcal{L}$ obtained by applying random rewiring with those for random link switching (compare Fig. 7.7A and Fig. 7.7B).

Additionally taking the effects of the nodes' spatial embedding into account, it is observed that GeoModel I generates random surrogates of all networks under study

for which the average path length $\mathcal{L}$ already becomes very close to its original values (Fig. 7.7C). However, while for the airline, Internet, world trade and Erdős-Rényi networks, the surrogates are also in good agreement with respect to deviations in the global clustering coefficient, one observes that GeoModel I still shows a positive bias of $\mathcal{C}$ for the remaining networks. Thus, for the aforementioned networks, the global link length distribution $F(l)$ already determines the expected value of the average path length $\mathcal{L}$, while the global clustering coefficient $\mathcal{C}$ is not yet explained sufficiently.

This mismatch is, however, to a large extent addressed by the usage of GeoModel II (Fig. 7.7D). All networks now display a reduction of the deviation in $\mathcal{C}$ as compared with the application of GeoModel I (compare Fig. 7.7C and Fig. 7.7D). This means that ultimately, in addition to the global link length distribution $F(l)$, the local link length distribution $F_i(l)$ predetermines in most cases and to a large extent the value of the global clustering coefficient $\mathcal{C}$.

In summary, one class of spatial networks (including the spatially embedded Erdős-Rényi, airline, world trade and Internet network) emerges for which random network models that do not take into account any spatial embedding of the nodes already generate surrogates with global clustering coefficients and average path lengths similar to those of the original networks. For a second class of networks (the power grid and random geometric graph as well as the interstate and urban road network) only taking the spatial structure of the original network explicitly into account in terms of GeoModel I and/or GeoModel II produces surrogates with global clustering coefficients and average path lengths comparable with those of the respective original networks. Remarkably, GeoModel I serves to reproduce well the average path lengths of the aforementioned networks, while only the application of GeoModel II produces network surrogates for which the global clustering coefficient becomes also close to the respective original network's value.

It should be emphasized, that the first class of networks, which includes the airline network, is generally non-planar (see Fig. 7.2B for a visual impression). In contrast, those networks where spatial embedding is found to have a larger influence on macroscopic properties are almost or even fully planar (as, e.g., the US interstate network in Fig. 7.2A). This hints to a probable relationship between the macroscopic properties studied in this work and the planarity of networks. An extension of the proposed models to also address these effects would help to further comprehend the mechanisms generating the observed network topologies. Note again that the models presented in this study are designed such that constraints are added to lower-order models in a consecutive manner (GeoModel I adds one additional constraint $\mathcal{E}_2$ to random link switching while GeoModel II adds one constraint $\mathcal{E}_3$ to GeoModel I). This nested model structure allows for a precise tracking of the effects that cause the surrogates' global network characteristics to converge into those of the original

network. However, it is not feasible to address the planarity constraint in a similar fashion. This constraint rather requests to also consider the actual geographical position of each node in the metric space which is not dealt with so far. In addition, further constraints would continuously reduce the set of possible surrogates that can be generated from a given network to an even lower size than GeoModel II does. Thus, in order to also address the issue of planarity a different approach of rewiring and network generation than the one presented here seems to be more appropriate. Future work should thus focus on this important issue and evaluate possible construction mechanisms for surrogate networks that preserve the planarity of a given network in addition to the spatial constraints discussed in this work.

Further evaluation of link length distributions

All results obtained in the previous sections indicate a classification of the networks under study into two different classes – (i) those where spatial embedding largely determines the networks' topology and (ii) those which are less affected by spatial constraints. This section now aims to further disentangle this typology based on the original networks' link length distributions. For this purpose, one first normalizes for each network the distances D_{ij} between all nodes i and j by the maximum possible distance $\max\{D_{ij}\}_{i,j}$ and thus obtains normalized distances $D'_{ij} \in [0, 1]$. From there it is possible to estimate the distribution $f(l')$ of normalized link-lengths $l' \in [0, 1]$ for each network using a Gaussian kernel with a bandwidth chosen according to Scott's Rule [Scott, 2015].

Now assume that $f(l')$ factorizes into two disjoint probability distributions: First, $f_{geo}(l')$ indicates the probability to find a distance in the interval $[l', l' + dl']$ between two randomly chosen nodes i and j that are not necessarily connected by a link, i.e., $f_{geo}(l')$ denotes the geographic inter-node distance distribution between all possible pairs of nodes in the network. Second, let $f_{int}(l')$ denote the *intrinsic* probability of the network to form a link with a length within the interval $[l', l' + dl']$ between two nodes i and j that display a corresponding geographical distance taken from the same interval.

The two probability densities $f(l')$ and $f_{geo}(l')$ can be easily estimated from the data. Recall that $f(l')$ denotes the distribution of all link lengths, while $f_{geo}(l')$ give the distribution of all inter-node distances. The quantity of interest in the scope of this work is the residual probability $f_{int}(l')$ which remains after factorizing out the general spatial constraints of the systems. From

$$f(l') = f_{geo}(l') f_{int}(l') \tag{7.16}$$

it is possible to immediately compute $f_{int}(l')$. The corresponding cumulative distribution functions $F_{int}(l')$ for all networks under study are displayed in Fig. 7.8A.

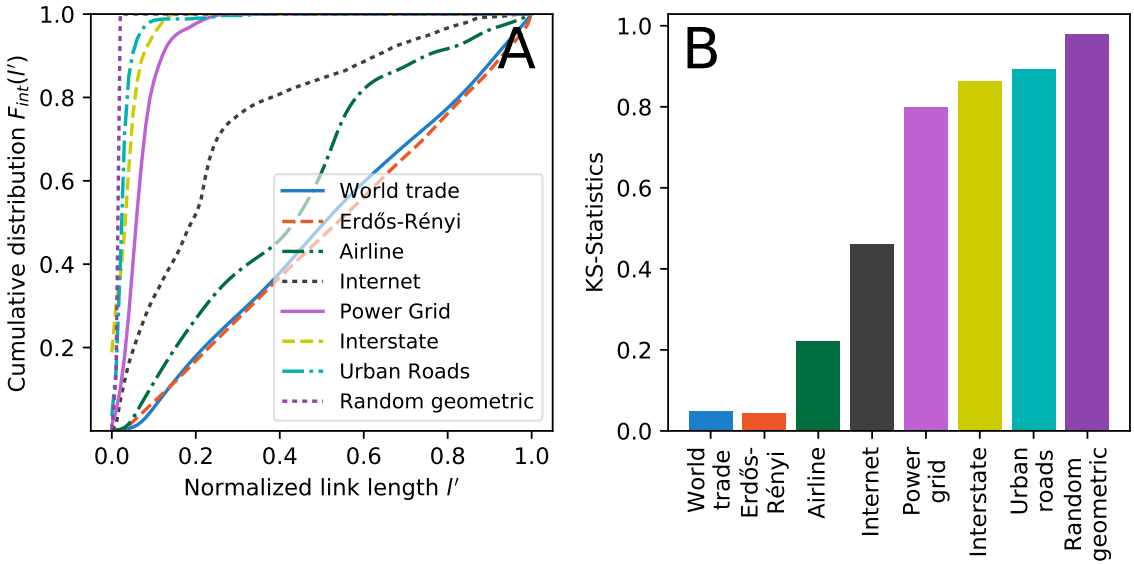

Figure 7.8.: (A) Cumulative intrinsic probability distribution $F_{int}(l')$ for the different networks under study. (B) KS statistics indicating the maximum distance between each cumulative probability distribution $F_{int}(l')$ and that of a random uniform distribution for each network under study sorted in ascending order (adapted from Wiedermann et al. [2016b]).

The four networks (power grid, interstate, urban roads and the random geometric graph) that were previously identified as being strongly influenced by the spatial embedding of the system display a strong increase in $F_{int}(l')$ for low values of l'. This signature is typical for exponential distributions and, thus, the intrinsic probability for links to be present in the network depends strongly on the inter-node distances even after accounting for all trivial geometric factors (Fig. 7.8A). In contrast, the world trade, airline and Erdős-Rényi network display a cumulative intrinsic link-length distribution $F_{int}(l')$ that increases almost linearly, a typical signature of a uniform distribution. Thus, for these networks, the actual linking probability that emerges after eliminating all trivial spatial effects is mainly random and does not depend much on the spatial embedding of the system (Fig. 7.8A).

The Internet presents a case slightly in between these two qualitatively different behaviors, exhibiting neither a strictly linear increase in $F_{int}(l')$ nor a sharp increase for small values of l'. This aligns well with the observation made from Fig. 7.7B, where the Internet sticks out in the sense that its average path length is already fully determined by the system's spatial embedding while its global clustering coefficient is not. This signature is unique to the Internet when compared to all networks under study.

A summary of these findings is shown in Fig. 7.8B which displays the KS dis-

tances for each of the intrinsic cumulative distribution functions $F_{int}(l')$ from a uniform distribution. One clearly notes the low values for the world trade, airline and Erdős-Rényi network, the intermediate value for the Internet and large values, i.e., large deviations from purely random behavior for the power grid, interstate, urban roads and the random geometric graph. Thus, the observed intrinsic link length distributions underline the findings from the previous sections and point again to a distinction of spatial networks into different distinct classes.

From a mechanistic point of view, it thus seems that there exists one class of networks for which global network characteristics can already be well reproduced by putting links between nodes according to a probability distribution that is purely determined by their distances. However, for a second class a second construction principle is superimposed that modulates this node-distance distribution such that short links are preferred over long-range connections.

Certainly, the assumption of independence between the intrinsic and geographic probability distributions $f_{int}(l')$ and $f_{geo}(l')$ is very strong and needs proof in the course of upcoming research. Specifically, this factorization assumes that nodes and links in a spatially embedded network are in a sense created independently from each other. This might to some extent be true for, e.g., the world trade network where first countries emerge and trade connections are then put between them, or the airline network, where already existing cities are in a second step connected by flights. In contrast, for systems such as road networks the creation of links representing streets and nodes representing intersections is naturally closely entangled and thus the above assumption certainly needs further evaluation. However, the findings from this section are mainly meant to present a first approximation of possible mechanisms behind the observed classification of spatial networks into different groups and suggest an interesting direction to further investigate the nature of spatially embedded networks. Future work should pick up these lines of thought to investigate more thoroughly the mechanisms that drive the emergence of links in those different types of networks.

7.6. Summary

This chapter introduced two novel models to generate random surrogates of a given spatial network that preserve either the global or the local distribution of link lengths between individual nodes and, hence, explicitly accounts for the embedding of the network in some metric space. The macroscopic properties of the resulting surrogates have been characterized by means of the global clustering coefficient and the average path length and were compared to those of the original networks from which the surrogates were constructed. For reference, iterative random rewiring and random link switching schemes have been utilized to produce random networks similar to

Erdős-Rényi networks and the configuration model, respectively.

It was found that for a certain class of spatial networks random link switching already produces surrogates of comparable macroscopic structure as the original network. Thus, for these networks the spatial embedding of the nodes and links is not crucial for explaining their corresponding macroscopic properties. In contrast, another class of networks was identified for which global clustering coefficients and average path lengths are only well reproduced when applying the newly introduced GeoModel I and/or GeoModel II that explicitly account for the spatial embedding of the nodes. Hence, for these networks information about their geometric properties is needed to sufficiently explain their macroscopic structure. For the latter class of networks, it was found that their average path length can already be well reproduced by GeoModel I, while only using GeoModel II enables to also capture the global clustering coefficient to a large extent. These findings align well with recent studies on the effect of the networks' spatial embedding on the small-world property of a system [Bialonski et al., 2010]. This work confirmed that the two quantities that are commonly assessed when determining whether a network displays the small-world property are in many cases to a large extent already predetermined by the spatial distances between its nodes.

In summary, the surrogate network models introduced in this work provide an important step in assessing whether and to which extent global characteristics of a complex network are already predetermined by statistics associated with the spatial embedding of its nodes and links. For future work, it will be of great importance to study in more detail which classes of networks are explicitly affected by the nodes' spatial embedding and which are already sufficiently quantified by some structural quantities such as the degree distribution. This issue might also imply to specifically impose constraints on the planarity of the surrogate networks corresponding to the original networks' topologies.

It was further observed that the optimal tolerance ϵ (the only parameter of the models introduced here) varies strongly depending on the specific networks under study. Therefore, the interplay between this parameter and the shape of the cumulative distribution function of link lengths as well as the number of nodes N_x and links M should be addressed in future research. This could help to directly estimate an optimal tolerance circumventing the need for evaluating individual KS statistics as applied in the present work. Additionally, it is of interest to extend the models presented in this work to also conserve degree correlations [Zamora-López et al., 2008] and to be also applicable to weighted networks, such as airline networks, where the weight of each link scales with the number of passengers on the corresponding connection [Barrat et al., 2004a]. In order to compare different types of networks, a thorough, concise and general definition of link weights must be found as otherwise no useful intercomparison of obtained weighted network characteristics

is possible (see also Sec. 7.1). In order to develop a corresponding framework, the present models could be combined with existing models for non-spatially embedded weighted networks [Ansmann and Lehnertz, 2011] that follow a similar strategy of constrained rewiring of a given network as the models presented in this work.

Chapter 8.

Discrimination of networks in the complexity-entropy plane

Up to now, all previous chapters have dealt with the assessment of functional or structural complex networks by means of analyses of their respective topological and/or spatial characteristics. However, even with the presented variety of characteristics at hand it still remains a subject of current research to appropriately quantify a network's actual *complexity* and correspondingly discriminate between different types of complex networks, like infrastructure or social networks, on such a basis. This chapter ultimately explores the possibility to classify complex networks by means of a statistical complexity measure that has formerly been successfully applied to distinguish different types of chaotic and stochastic time series. It is composed of a network's averaged per-node entropic measure characterizing the network's information content and the associated Jenson-Shannon divergence as a measure of disequilibrium. In a first application, the proposed formalism is applied to 29 real-world networks that are discriminated by context into the four types of social animal, social affiliation, transportation and connectome networks. In most cases the different types occupy distinct areas in the resulting complexity entropy-plane. Thus, the formalism introduced here naturally distinguishes between different types of systems under study. In particular and within the proposed framework, connectome networks exhibit among the highest complexity while, e.g, transportation and infrastructure networks display significantly lower values. A second application then links back to the study of functional networks as presented in Part I and shows that the proposed framework is useful to objectively construct threshold-based networks, such as functional climate networks, by choosing the threshold in a way that the statistical network complexity is maximized.

The remainder of this chapter is organized as follows. Section 8.1 introduces a summary of previous attempts to study and quantify a network's complexity and outlines the approach proposed in this chapter based on these former assessments. After some preliminary observations in Sec. 8.2, Section 8.3 introduces the specific notion of entropy that forms the basis of the framework that is proposed here. From

there, Sec. 8.4 ultimately derives a network complexity measured which is then applied to classify the 29 real-world networks of different categories as they are presented in Sec. 8.5. Section 8.6 introduces (in addition to functional climate networks) a second type of threshold based networks denoted as *recurrence networks* for which optimal thresholds are estimated by means of the methodology introduced here. All corresponding results are presented in Sec. 8.7 and the chapter is ultimately concluded with a summary and an outlook in Sec. 8.8.

This chapter is primarily based on the study that is presented in P13 [Wiedermann et al., 2017a].

8.1. Theoretical background

As outlined in Chap. 2, 4, 6, and 7, the structure of complex networks may be well quantified by a set of topological [Albert and Barabási, 2002] (Chap. 4 and 6) or spatial [Barthélemy, 2011] (Chap. 7) characteristics which measure certain properties of either distinct nodes (local characteristics) or the entire network itself (global characteristics). However, while the large variety of present metrics allows for a quantification of a network's particular macroscopic and microscopic structure, it still remains a subject of current research to (i) assess the actual *complexity* of a network based on these sets of characteristics [Anand and Bianconi, 2009] and (ii) to determine distinct sets of properties for certain classes of networks, such as infrastructure or social networks, in order to objectively and comprehensively distinguish between them. While there exists a variety of such complexity measures [Bonchev and Buck, 2005], most of them are tailored to specific applications and have so far not been successfully applied to intercompare different types or classes of networks as in this respect they often lack a meaningful interpretation [Dehmer et al., 2009].

Contributing to the above issues, this chapter introduces a two-dimensional metric based on an entropic and an adjoint statistical complexity measure to distinguish different types of complex networks. This approach was originally introduced to distinguish chaotic from stochastic systems in time series analysis[Martin et al., 2006, Rosso et al., 2007] and has been successfully applied to study, e.g., ordinal patterns in daily stream time series of river runoff [Lange et al., 2013]. Its purpose is to assign each system under study a position in a two dimensional space spanned by an entropy and a statistical complexity measure, the latter being a product of entropy and Jenson-Shannon divergence with respect to a uniform distribution.

Here, this concept is transformed from time series to the case of complex networks by accordingly redefining the above entropy and statistical complexity. Various definitions of network entropies or, more specifically, the underlying probability distributions have already been proposed [Dehmer et al., 2012]. They may for example

be computed in terms of the network's topological information content [Rashevsky, 1955] or, quite commonly, its degree distribution [Bonchev, 1983, Cancho and Solé, 2003, Wang et al., 2006]. Further definitions of entropy are based on the assessment of network ensembles or randomized correspondents thereof [Anand and Bianconi, 2009, Bianconi, 2008] However, particularly entropy measures that are based on the degree-distribution alone have been shown to have little discriminative power when applied to a heterogeneous set of networks [Dehmer et al., 2012]. It is in contrast rather advisable to rely on local per-node definitions of network entropies [Dehmer et al., 2012, Konstantinova et al., 2003]. Therefore, this work relies on one recently proposed candidate for such a node-wise definition of entropy that is based on the probability of a random walker to jump from a specific node to its neighbors in the network [Small, 2013]. This notion of entropy is closely related to random walks which themselves are in their application and interpretation closely related with the assessment of a networks' navigability and thus, complexity [Amaral and Ottino, 2004, Barthélemy, 2011, White and Houseman, 2002]. Its specific details are discussed in Sec. 8.3.

8.2. Preliminaries

Analogously to the concept of the *complexity-entropy plane* in nonlinear time series analysis which has been utilized to discriminate between different types of time series generated by stochastic and deterministic chaotic processes [Martin et al., 2006, Rosso et al., 2007], this chapters aims to characterize a set of complex networks by means of its average per-node Shannon entropy H and a statistical complexity measure C. Thereby two notions are used which are related to the complexity of a physical system, namely its *information content* and its state of *disequilibrium* [Calbet and López-Ruiz, 2001, Lamberti et al., 2004, López-Ruiz et al., 1995]. In particular, the information content of the network is linked to the entropy H and the disequilibrium is expressed in terms of the network's Jenson-Shannon divergence Q with respect to an appropriately chosen reference state.

Before going to the case of complex networks, an analogy to classical statistical physics is discussed in terms of the two most extreme cases of complexity one might consider, namely the crystal as well as the ideal gas displaying a large and a low degree of order, respectively [Anderson, 1991, López-Ruiz et al., 1995, Parisi, 1993].

It is easily deductible that due to its regular structure, the crystal usually contains low or almost zero information and, hence, $H \to 0$. In contrast to this, the ideal gas (due to its disorder) contains a large amount of information, implying $H \gg 0$. Further, it is observed, that the perfect crystal displays among the highest disequilibrium ($Q \gg 0$), i.e., a large degree of order, while the ideal gas displays the exact opposite ($Q \to 0$). As both measures, H and Q, usually increase or decrease mono-

tonically [López-Ruiz et al., 1995] their individual assessment can not be directly associated with a notion of complexity. Therefore, this work follows the proposed framework by López-Ruiz et al. [1995] and derives a measure of statistical complexity C as the product of both, information content (e.g, Shannon entropy H) and disequilibrium (e.g, Jenson-Shannon divergence Q). In the following the notion of information content and disequilibrium is transferred to the case of complex networks and corresponding terms for the entropy H and the statistical complexity C are derived. These metrics then ultimately form the two-dimensional complexity-entropy plane.

8.3. Network entropy

Generally, the classical Shannon entropy H for a discrete probability distribution f is given by

$$H(f) = -\sum_i p_i \log p_i. \tag{8.1}$$

Here, p_i denotes the probability of occurrence for a given state i. Since averaged per-node entropies have been shown to generally serve as a good choice for discriminating between different types of networks [Dehmer et al., 2012], this work chooses one specific definition of entropy that is based on the assessment of probabilities to jump between nodes when randomly traveling through the network. This definition has formerly been successfully applied to the study of complex networks constructed from univariate time series [Small, 2013]. In particular, the entropy H_i for each of the N_x nodes i in the network is computed based on the distribution f_i with entries $p_{i\to j}$ that give the uniformly distributed probability to jump from node i to node j in exactly one step if a link exists between them. Thus, the corresponding random walk is formulated analogously to its application in computing the recently proposed *random walk betweenness* [Newman, 2005]. For an unweighted and undirected network the corresponding probabilities $p_{i\to j}$ are then given by

$$p_{i\to j} = \frac{A_{ij}}{k_i} \in \{0, 1/k_i\} \;\forall\; k_i > 0 \tag{8.2}$$

with $\sum_j p_{i\to j} = 1$. The node entropy then reads,

$$H_i(f_i) = -\sum_{j=1} p_{i\to j} \log p_{i\to j} = -\sum_j \frac{A_{ij}}{k_i} \log \frac{A_{ij}}{k_i} = \log k_i \;\forall\; k_i > 0. \tag{8.3}$$

In case of an isolated node i with $k_i = 0$ the entropy is set as $H_i(f_i) = 0$. Ultimately, the average normalized entropy taken over all nodes i is referred to as the network

entropy,

$$H(f) = \frac{1}{N_x \log(N_x - 1)} \sum_i \log k_i \in [0, 1]. \tag{8.4}$$

Note that $H(f)$ assumes the network under study to display no trivial self-loops which manifests in the specific normalization constant $\log(N_x - 1)$. This specific definition of entropy is in congruence to some magnitude-based information indices [Bonchev, 1983] that measure tendencies for complex networks to form branches. In particular, the above measure quantifies the heterogeneity in the network's degree distribution in a sense that nodes with low degree, i.e, peripheral nodes lower the overall network entropy $H(f)$ while high degree-hubs cause its increase. Furthermore, it follows from the corresponding series expansion of $H(f)$ around the average degree K, that the present definition of $H(f)$ incorporates not only average statistics of the network's degree distribution but implicitly also accounts for its higher moments, such as the variance,

$$H(f) = \frac{\langle \log k_i \rangle_i}{\log(N_x - 1)} \approx \frac{\log K}{\log(N_x - 1)} - \frac{\langle (k_i - K)^2 \rangle_i}{2 \log(N_x - 1) K^2}. \tag{8.5}$$

The entropy $H(f)$ can also be interpreted with respect to the underlying formulation of the random walk. In the limiting case of a fully connected network the probability to jump between nodes is given as $p_{i \to j} = \frac{1}{N_x - 1} \; \forall \; i \neq j$. Thus, the walk becomes fully random in a sense that no node j is excluded as a possible candidate for the walker to jump to. This case is directly related to the notion of the ideal gas outlined above where all micro states are equally probable and, thus, the entropy is maximized. Analogously, for the fully connected network all walks through the network of arbitrary length are equally probable, too. Consequently, the entropy $H(f)$ is also maximized and reads $H(f) = 1$.

In turn, for a sparsely connected network the jumps of the walker become more deterministic and in the limiting case of, e.g., node i only having one neighbor n, its associated traverse probabilities approach $p_{i \to j} = \delta_{jn}$ (with δ_{jn} being Kroneckers Delta). In this case, the walker only has one option for jumping to a neighboring node of i. Consequently, the entropy is lowered and for sufficiently sparse networks approaches $H(f) \to 0$. Again, this case may be interpreted in analogy to a regular crystal that displays perfect order as well as a deterministic structure and, hence, has a low information content. In summary, the above definition of $H(f)$ as a measure of regularity or order in the network under study is thus interpreted with respect to a network's navigability that is measured in terms of a random walk.

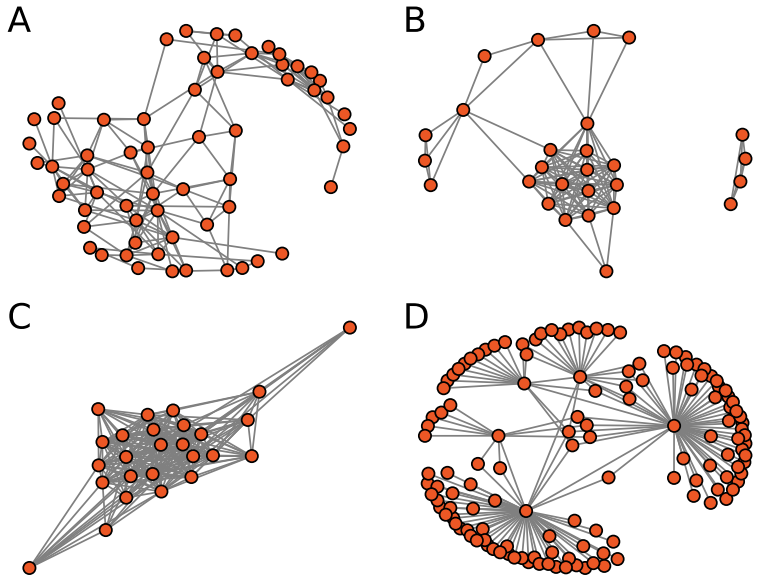

Figure 8.1.: Selection of four out of the 29 networks investigated in this study: Dolphin (A), zebra (B) and bison (C) animal social networks as well as the network of American revolutionary groups (D) (adapted from Wiedermann et al. [2017a]).

8.4. Statistical complexity

This work aims to express statistical complexity or non-triviality in terms of a system's disequilibrium and information content [López-Ruiz et al., 1995], with the latter being defined as the network entropy $H(f)$ that is introduced above. Along the lines of common statistical mechanics, disequilibrium is measured here in terms of the Jenson-Shannon divergence [Martin et al., 2003]

$$Q_i(f_i, f_{i,e}) = Q_0\{H_i(0.5[f_i + f_{e,i}]) - 0.5[H_i(f_i) + H_i(f_{e,i})]\}, \tag{8.6}$$

with $Q_0 = 1/\log 2$ to ensure $Q_i \in [0, 1]$. Here, the probability distribution f_i with entries as introduced in Eq. (8.2) again denotes the probabilities to jump between neighboring nodes i and j when randomly traveling through the network. The distribution $f_{e,i}$ denotes the same, but for an appropriately chosen reference or equilibrium state, i.e., network. Note, that defining the equilibrium state of a network is a highly non-trivial task that is often achieved by fitting the network under study to a certain network model using numerical variational techniques in order to minimize or maximize a target or cost function [Dafermos and Nagurney, 1984, Giannessi and Maugeri, 1995]. However, in order to demonstrate the applicability of the proposed approach, it is in analogy with previous works by Martin et al. [2003] assumed that a system is in equilibrium if its state corresponds to the fully randomized one. For the case of complex networks, the equilibrium state would than be a corresponding

Erdős-Rényi [Erdös and Rényi, 1960] random network.

Analogously to the network entropy, the Jenson-Shannon divergence Q of the entire network is again computed as the arithmetic mean of all per-node values Q_i,

$$Q(f, f_e) = \frac{1}{N_x} \sum_i Q_i(f_i, f_{i,e}). \tag{8.7}$$

As for the case of the network entropy $H(f)$, the analogy between generic physical systems (perfect crystal and ideal gas) is apparent. Again, the fully connected network corresponds to the case of an ideal gas with minimum disequilibrium as any appropriately chosen reference network should be fully connected as well, which implies $f = f_e$ and, thus, $Q = 0$. In contrast, a randomly chosen reference to a sparsely connected network most certainly displays a different microscopical structure and, hence, the probabilities for jumping between nodes f and f_e also differ, yielding a high disequilibrium $Q \gg 0$.

With the above observations in mind, it is demanded based on common sense that neither the fully connected nor the very sparsely connected (or almost empty) network should be attributed a large complexity. Hence, neither a measure of *information* $H(f)$ nor *disequilibrium* $Q(f, f_e)$ alone may serve as an appropriate quantifier of statistical complexity. Addressing this particular issue, a measure of complexity C has been proposed that is based on a product of the two quantities [López-Ruiz et al., 1995, Rosso et al., 2007],

$$C(f, f_e) = Q(f, f_e)H(f) \in [0, 1]. \tag{8.8}$$

This measure intuitively exhibits the required asymptotic properties, such that for the limiting case $H(f) = 0$, it follows that $C(f, f_e) = 0$. Analogously, $H(f) = 1$ is only achieved for a fully connected network which implies $f = f_e$ and $Q(f, f_e) = 0$, which also yields $C(f) = 0$. For all cases $0 < H(f) < 1$, the statistical complexity $C(f)$ has a possible upper bound that is determined by $H(f)$. However, its analytical expression is highly non-trivial and has so far only been obtained for a binary state probability distribution [Lamberti et al., 2004].

As the equilibrium or reference state of a given network is defined as its fully randomized counterpart, C may thus be interpreted as an indicator of statistical independence from a corresponding random state. To account for the stochasticity of this model, an ensemble of $N_r = 100$ independent Erdős-Rényi networks with the same number of nodes N_x and linking probability $d = \sum_i k_i/(N_x(N_x - 1))$ is computed as the reference for each network under study. From there ensemble average quantities of Q_i and Q are computed from the resulting probability distributions $f_{i,e}$.

Ultimately note, that a variety of further measures has been developed that similarly aim to quantify complexity in dynamical systems [Wackerbauer et al., 1994]. However, most of these measures are more tailored to other applications, such as

Name	Category	N_x	d	H	C
Sheep	Soc. Ani.	28	0.622	0.825	0.326 ± 0.022
Rhesus	Soc. Ani.	16	0.575	0.753	0.341 ± 0.035
Kangaroo	Soc. Ani.	17	0.669	0.782	0.285 ± 0.024
Mac	Soc. Ani.	62	0.617	0.874	0.339 ± 0.009
Bison	Soc. Ani.	26	0.683	0.855	0.281 ± 0.019
Zebra	Soc. Ani.	27	0.316	0.571	0.409 ± 0.024
Cattle	Soc. Ani.	28	0.542	0.772	0.372 ± 0.018
Dolphins	Soc. Ani.	62	0.084	0.342	0.317 ± 0.006
Autobahn	Transp.	1,168	0.002	0.099	0.099 ± 0.000
USairport500	Transp.	500	0.024	0.250	0.246 ± 0.001
USairport 2010	Transp.	1,574	0.014	0.248	0.246 ± 0.000
Openflights	Transp.	2,939	0.004	0.173	0.173 ± 0.000
Rome99	Transp.	3,353	0.001	0.121	0.121 ± 0.000
South-Africa	Soc. Affil.	6	0.633	0.601	0.306 ± 0.062
American Revolution	Soc. Affil.	136	0.017	0.043	0.043 ± 0.001
Club-Membership	Soc. Affil.	25	0.305	0.576	0.415 ± 0.021
Corporate-Leadership	Soc. Affil.	24	0.322	0.590	0.417 ± 0.026
Rhesus Brain 1	Connectome	242	0.105	0.523	0.474 ± 0.002
Rhesus Brain 2	Connectome	91	0.142	0.452	0.401 ± 0.006
Mouse Retina 1	Connectome	1,076	0.157	0.693	0.594 ± 0.001
Mixed Species Brain 1	Connectome	65	0.351	0.717	0.478 ± 0.012
Rhesus Cerebral Cortex 1	Connectome	91	0.342	0.710	0.485 ± 0.007
C Elegans Neural Male 1	Connectome	269	0.081	0.486	0.451 ± 0.002
Rattus Norvegicus Brain 3	Connectome	493	0.214	0.684	0.558 ± 0.001
Rhesus Interarea Cortical 2	Connectome	93	0.529	0.822	0.413 ± 0.006
Rattus Norvegicus Brain 2	Connectome	502	0.196	0.666	0.556 ± 0.001
Rattus Norvegicus Brain 1	Connectome	503	0.182	0.653	0.553 ± 0.001
Mouse Brain 1	Connectome	213	0.716	0.934	0.272 ± 0.003
C Elegans Herm Pharynx 1	Connectome	279	0.059	0.460	0.436 ± 0.002

Table 8.1.: Overview of the networks evaluated in this study together with their respective number of nodes N_x and link density d as well as entropy H and statistical complexity C computed over an ensemble of $N_r = 100$ reference networks. The provided estimate of the error in C denotes one standard deviation. Categories (Soc. Ani. = Social Animal, Transp. = Transportation and Soc. Affil. = Social Affiliation) have been assigned according to their classification in the Colorado Index of Complex Networks (ICON) (`https://icon.colorado.edu/`).

bifurcation analysis or the detection of chaos transitions. Thus this chapter focuses on the statistical complexity measure as introduced above.

8.5. Real-world networks

This work studies the entropy H and statistical complexity C of 29 real world networks for which certain types are assigned according to their sub-domains in the Colorado Index of Complex Networks (ICON) (`https://icon.colorado.edu/`). Specifically, eight networks that represent social networks among different species of animals, five transportation networks, four networks representing affiliations between people or corporations and twelve connectome networks for different species are obtained. In order to make the results comparable, all networks under study are treated unweighted, undirected and without self-loops. The networks under study, together with their assigned type, number of nodes N_x and link density d are summarized in Tab. 8.1 and the corresponding references to each network are given in Tab. A.1 in the appendix. Visual representations of the topological structure of four of the 29 networks are shown in Fig. 8.1. The network parameters N_x and d will be used to compute corresponding reference networks as outlined in Sec. 8.4.

8.6. Threshold-based networks

In addition to (structural) real-world networks, this works also aims to illustrate the usefulness of the statistical complexity C as a measure to objectively construct threshold-based networks. Besides functional climate networks (Part. I), a second example of such networks are so-called recurrence networks [Donner et al., 2010, Lange and Boese, 2015]. Here, nodes usually indicate different temporal sampling points in a time series and links are put between them if they are closely neighbored in phase space. The analysis of recurrence networks has in the past been successfully applied to, e.g., detect transitions in paleo-climate variability with respect to human evolution [Donges et al., 2011c] or to classify weak and strong phases in the evolution of the East Asian summer monsoon and the Indonesian-Australian summer monsoon [Eroglu et al., 2016]. As one of the most prototypical examples, this work studies a recurrence network that is constructed from the three-dimensional Rössler system given by

$$\frac{dx}{dt} = -y - z \tag{8.9}$$

$$\frac{dy}{dt} = x + ay \tag{8.10}$$

$$\frac{dz}{dt} = b + z(x - c). \tag{8.11}$$

Here, $a = b = 0.2$ and $c = 5.7$ are set according to the original study of this system [Rössler, 1976]. In the past, recurrence networks have been shown to capture essential information on the phase space structure of the dynamical system under

study and thus serve as a good (or even equivalent) representation of the system's trajectory [Donges et al., 2012, Donner et al., 2011]. Each node i in the network represents a point $\mathbf{x}_i = (x(t_i), y(t_i), z(t_i))$ on the system's trajectory at randomly chosen times $t_i \in [100, 1000]$, where $t_i \geq 100$ ensures that for the choice of initial values $x(0) = y(0) = z(0) = 1$ the system has converged onto the chaotic attractor. The entries D_{ij} of the distance matrix $\mathbf{D}$ are then given by the Euclidean distances between points $\mathbf{x}_i$ and $\mathbf{x}_j$ [Donner et al., 2010]. From $\mathbf{D}$, a corresponding recurrence matrix $\mathbf{R}$ with entries R_{ij} is constructed by choosing a recurrence threshold T such that

$$R_{ij} = \Theta(T - D_{ij}) \tag{8.12}$$

where $\Theta(\cdot)$ again denotes the Heaviside function. $\mathbf{R}$ is now interpreted as the adjacency matrix $\mathbf{A}$ of a spatial recurrence network such that $A_{ij} = R_{ij} - \delta_{ij}$. Hence, only distances between points that are smaller than a critical distance T are connected in the resulting network. In full analogy to functional climate networks, the threshold T is usually chosen such that a desired link density, also called *recurrence rate*, d is obtained. This work illustrates in Sec. 8.7 that the notion of statistical network complexity C can be utilized as a complement to the above approach by choosing T such that C is maximized.

To illustrate the applicability of statistical complexity as a metric to objectively chose thresholds T also for functional climate networks, two example networks are computed from (i) monthly averaged surface air temperature and (ii) monthly averaged sea level pressure time series provided by the NCEP/NCAR 40-year reanalysis project [Kalnay et al., 1996] of the National Center of Oceanic and Atmospheric Research. As analogue to the studies presented in Part I the data is pre-processed according to Sec. 3.1 and the similarity matrix $\mathbf{S}$ is again composed of the absolute values of all pairwise Pearson correlation coefficients at zero lag (Sec. 3.2).

8.7. Results

The first application now studies the numerical values of entropy H and complexity C for the different real world networks as presented in Sec. 8.5. A second use case illustrates the application of statistical complexity to objectively determine appropriate thresholds for the construction of functional networks.

Real world networks

Figure 8.2 displays the entropy H and statistical complexity C of all real world networks under study averaged over an ensemble of $N_r = 100$ Erdős-Rényi reference networks (the average numerical values of H and C are also presented in Tab. 8.1).

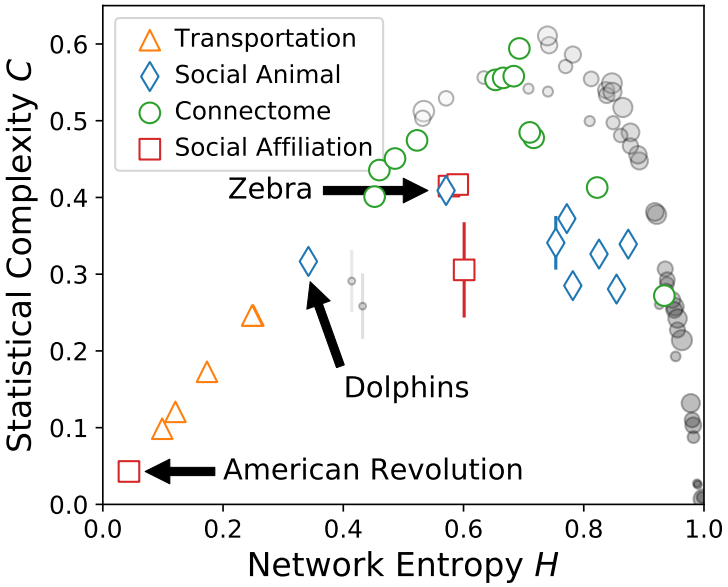

Figure 8.2.: Mapping of real-world networks in the complexity-entropy plane. Additionally, gray scatter show the results for 50 different realizations of an Erdős-Rényi random network. Here, the size and transparency denotes the uniformly at random drawn number of nodes N_x from the interval $[10, 1000]$ and linking probability d from the interval $[0, 1]$, respectively. Error bars indicate one standard deviation of statistical complexity taken over an ensemble of $N_r = 100$ random Erdős-Rényi reference networks and are shown were their size exceeds that of the corresponding symbol (adapted from Wiedermann et al. [2017a]).

For reference, the complexity and entropy of a set of 50 Erdős-Rényi networks with the number of nodes N_x and linking probability d drawn uniformly at random from the intervals $[10, 1000]$ and $(0, 1]$ is computed and presented as well.

The first major observation is that different types of networks under study generally occupy distinct areas in the complexity-entropy plane (Fig. 8.2). While connectome networks show among the highest values of C, intermediate values are observed for both types of social networks and the lowest values emerge for the transportation networks. Additionally, the latter also exhibit among the lowest values of entropy H. Notable exceptions are the social networks of dolphins and zebras, which in contrast to most of the other animal networks display a unique community structure (see Fig. 8.1A,B for a visual representation). Specifically, the dolphin network (Fig. 8.1A) is characterized by two distinct communities that are connected only via few nodes while the zebra network (Fig. 8.1B) is composed of one large almost fully connected community containing roughly half of the nodes and at least two further distinct communities with only few nodes that hardly connect with the main herd. In contrast, all the other animal networks (see Fig. 8.1C for a representative example) generally display a similar structure with only one densely connected community. Another

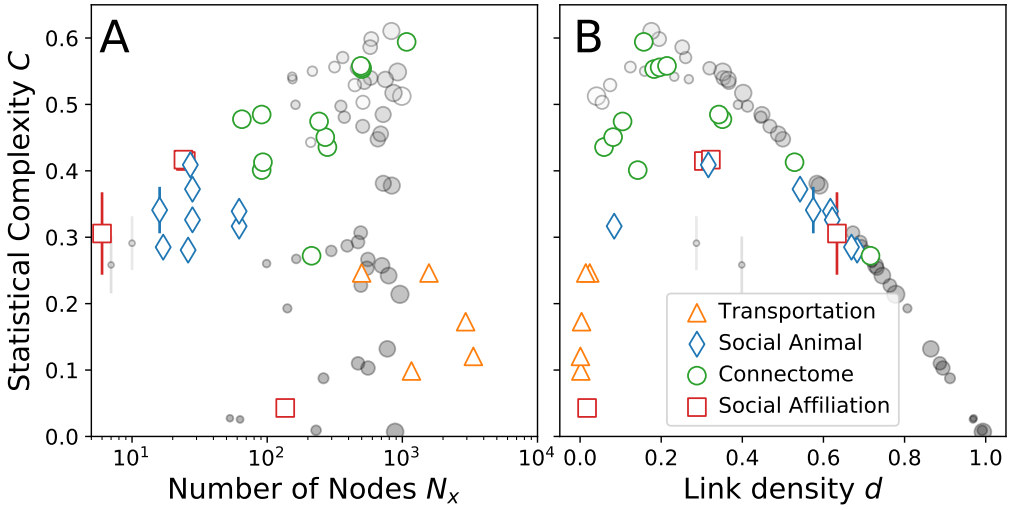

Figure 8.3.: (A) Dependence of the statistical complexity C on the number of nodes N_x in each network under study. (B) The same as in (A) for the link density d. Gray scatter show the results for 50 different realizations of an Erdős-Rényi random network as in Fig. 8.2. Error bars indicate one standard deviation of statistical complexity taken over an ensemble of $N_r = 100$ random Erdős-Rényi reference networks and are shown were their size exceeds that of the corresponding symbol (adapted from Wiedermann et al. [2017a]).

outstanding exception is the network of American revolutionary groups (Fig. 8.1D), which due to its distinct hierarchical community structure displays very low values of entropy and complexity. Despite the few outliers, these first observations imply that the complexity-entropy plane generally distinguishes well between different types of networks solely based in their specific and distinct topology.

The random Erdős-Rényi networks (gray scatter in Figs. 8.2 and 8.3) show in many cases a higher statistical complexity than real world networks. In fact, their values roughly seem to determine an upper bound of C for each possible value of H (Fig. 8.2). This behavior is expected, as the two networks that are compared in the Jenson-Shannon divergence $Q(f, f_e)$, the Erdős-Rényi network under study and a random reference network, are statistically fully independent by construction and, more importantly, therefore less or equally less statistically dependent than any real world complex network in comparison with a random reference network. However, this property only seems to hold for sufficiently large networks (Fig. 8.2).

Since the topological characteristics of the Erdős-Rényi network only depend on the given number of nodes N_x and linking probability or link density d, the dependence of C on both parameters is also examined individually. Figure 8.3A shows the values of statistical complexity C as a function of the number of nodes N_x in each network which displays no clear dependence between the two variables. In contrast, a possible dependence between link density d and statistical complexity C is observed

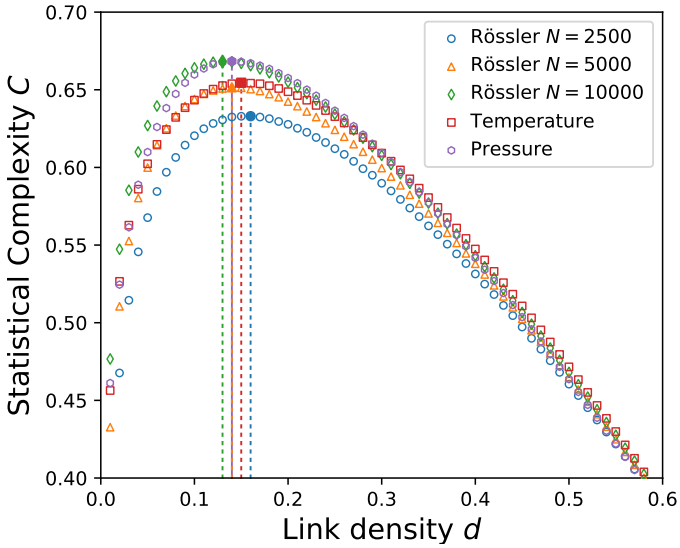

Figure 8.4.: Statistical complexity C depending on the threshold-based networks' link densities d for three recurrence networks with different numbers of nodes N_x constructed from the Rössler system and two functional climate networks representing surface air temperature and sea level pressure variabilities. Filled symbols denote the maximum value of statistical complexity for each network. Dotted lines indicate the corresponding link density d_{max}, which maximizes the statistical complexity. No error bars are shown as the standard deviation of C taken over all $N_r = 100$ reference networks is smaller than the size of the symbols (adapted from Wiedermann et al. [2017a]).

(Fig. 8.3B). Still, note that networks with highly dissimilar link densities d may exhibit similar statistical complexity (Fig. 8.3B).

Furthermore, the quantitatively similar functional dependencies between H and C (Fig. 8.2) as well as d and C imply an expected functional dependence between H and d. However, the H-C plane is a much better choice for categorizing networks than the d-C plane since the entropy H captures all moments in the degree distribution of a given network (as can be seen from the series expansion of $\sum_i \log k_i$ in Eq. (8.5)), while d only captures its first moment.

Threshold-based networks

In a second application the focus is turned to the threshold-based networks introduced in Sec. 8.6. Figure 8.4 shows the statistical complexity C of three recurrence networks with $N_x = 2,500$, $N_x = 5,000$, and $N_x = 10,000$ nodes obtained from the Rössler system (Eqs. (8.9)-(8.11)) depending on the link density d that is applied to obtain the recurrence matrix $\mathbf{R}$ (Eq. (8.12)). For all cases, C is computed as the average statistical complexity taken over an ensemble of $N_r = 100$ reference networks.

Note, that C increases with increasing d until a maximum is reached at $d_{max} = 0.16$, $d_{max} = 0.14$ and $d_{max} = 0.13$ for $N_x = 2,500$, $N_x = 5,000$ and $N_x = 10,000$, respectively (Fig. 8.4). For values $d > d_{max}$ the statistical complexity decreases monotonically and approaches $C = 0$ for $d = 1$ which has been checked numerically. Usually, when constructing recurrence networks a link density of order $\mathcal{O}(10^{-2})$ is chosen heuristically, even though it was suggested that with such comparably low choices of recurrence rates possibly significant interdependencies between nodes might be suppressed [Donges et al., 2011a, Donner et al., 2010, Marwan et al., 2009]. The results presented here indicate that opting for slightly larger choices of recurrence rates and corresponding thresholds T, might yield a recurrence network with higher statistical complexity thus, a larger degree of non-trivial structure than the ones that were previously typically studied. The assessment of statistical complexity could help to more objectively choose thresholds for the construction of such networks and complements existing approaches based on, e.g., the assessment of the recurrence network's percolation threshold [Donges et al., 2012, Eroglu et al., 2014, Jacob et al., 2016].

For the functional networks that are constructed from climate time series across the globe at different link densities d, one finds that the link densities that are maximizing the statistical complexity are $d_{max} = 0.15$ and $d_{max} = 0.14$ for the temperature and pressure field, respectively (Fig. 8.4). As for the recurrence network studied above, these values are again considerably larger then the usually employed link densities of order $\mathcal{O}(10^{-2})$ [Donges et al., 2009a,b] which have also been utilized for the analyses presented in Part I of this work. However, as seen in, e.g., Sec. 6.4 these usually employed small choices of link density and the corresponding high threshold may suppress statistically significant signals of comparably lower similarity values. Thus, future work on climate network analysis could apply the complexity-based formalism to determine a more objectively chosen threshold than in previous studies.

Even though this section only presents two different use cases as examples, the proposed framework may be applicable to any kind of functional network that is constructed from some pairwise functional interdependencies between nodes, including neural [Bota and Swanson, 2007, Bullmore and Sporns, 2009] or economic networks [Maluck and Donner, 2015]. Furthermore, this framework might also be applicable to networks constructed from non-pairwise interdependencies that are investigated in, e.g, causal effect networks [Kretschmer et al., 2016, Runge et al., 2015].

8.8. Summary

This chapter presented a methodology to categorize complex networks by means of an entropy measure and an estimator of statistical complexity. In particular, this method computes for each network under study an average per-node Shannon entropy that is based on probabilities to randomly jump between neighboring nodes in

the network. From this, a network's statistical complexity is estimated by computing the Jenson-Shannon divergence between a given network and a set of corresponding Erdős-Rényi random networks. It is observed that networks of different types, such as social or infrastructure networks, generally occupy distinct regions in the two-dimensional complexity-entropy plane and the proposed framework thus discriminates well between them. Moreover, it is observed that connectome networks are among the statistically most complex ones while infrastructure networks generally display a lower complexity. These properties might intuitively be expected when considering the term complexity with respect to real world structures and the associated functions thereof.

It was further shown in a second application that the notion of statistical complexity can be applied to objectively estimate thresholds for the construction of functional networks, such that a network's statistical complexity is maximized and, hence, contains most non-trivial information.

This chapter demonstrates possible scenarios for applying the proposed methodology. Future work should investigate in more detail the discriminating power of the statistical complexity for a broader set of real world complex networks. In particular, as within the framework presented here connectome networks are among the most complex ones, the interplay between the statistical complexity and the complexity of structure-function relations in such networks should be investigated in more detail in future work [Zhou et al., 2007]. Additionally, the framework should be generalized to the case of weighted and/or directed networks. For this purpose, more emphasis must be put into the definition of the reference networks, which for now have been assumed to just be a randomized correspondent of the specific network under study.

In general, the framework presented here expands the understanding of complex topological structures and helps to quantify varying degrees of complexity in various systems. As such, the approach should be useful for many disciplines of (applied) complex network science, such as neuro-, social and climate science.

This chapter closes the second part of this work that was devoted to the assessment and corresponding classification of real-world (structural) networks. The following final chapter concludes this work and summarizes all findings presented in part I and part II. Further, the resulting implications and outlines for future research are discussed.

Chapter 9.

Conclusion

The present work has primarily been devoted to study the application of complex networks as a tool to discriminate between different spatial and temporal stages within a complex dynamical system as well as to differentiate a larger set of given networks by means of certain network characteristics. Specifically, the work was split in two major parts. Part I studied the application of complex network theory to that of functional, in the present case, climate networks which proved useful to either discriminate different stages in the temporal evolution of the global climate system as well as to differentiate between different characteristics of interactions between two coupled climatic subsystems.

The respective Pt. I of this work opened with Chap. 4 which specifically addressed the ongoing, yet pressing, issue of reliably and objectively discriminating the two previously reported different phases of El Niño, the East Pacific (EP) and the Central Pacific (CP) kind. As both of these types have been shown to display differential imprints on the global climate system, a proper discrimination is key to reliably anticipate related impacts of either of them. In the past, such discriminations have been performed by assessing appropriately averaged sea surface temperature anomalies in the Equatorial Pacific or related linear statistic thereof, such as empirical orthogonal function analysis. However, it has been shown here that, when compiling a review of the recent literature, a majority of previous classifications is not performed in agreement. Contributing to this issue, this work proposed an index based on the assessment of the corresponding evolving functional climate network's transitivity with respect to an appropriately chosen long-term baseline. It has been shown that the index, in contrast to classically employed frameworks, provides a distinct and sharp discrimination between the two types. Specifically, all classifications that were in the past proposed in mutual agreement between all reviewed references could be confirmed. In addition, a consistent classification has been given for those periods were former works yielded ambiguous results. Furthermore, the index allows to perform a similar classification into two types for La Niña periods as well, hence contributing to another yet ongoing debate on the possibility of performing such a discrimination for the negative phase of the El Niño Southern Oscillation (ENSO). As specifically

for the latter comparable past works are scarce, future research should further evaluate the discrimination of La Niña periods proposed by the network-based method. On a more general matter, future work should also investigate more thoroughly the spatial distribution of links in the evolving climate network during different ENSO stages to gain a more systematic understanding of the physical mechanisms behind the observed differences in transitivity. As the proposed framework is independent of any manual thresholding and, hence, being automated and objective, it allows for a systematic evaluation of climate model simulations and could also be used to investigate potential changes in the projected frequency of the two ENSO flavors in the future, e.g., due to anthropogenic global warming.

With the network based discrimination of ENSO phases at hand, Chap. 5 moved towards an assessment of the different impacts of either of the four types, Central and Eastern Pacific type of El Niño and La Niña, on global extreme precipitation signals as one of the most significantly affected climatic variables related with ENSO variability. As the transitivity index, in contrast to classical indices of ENSO variability and classifications thereof, proved to already display a sharp spike-like signal it immediately qualifies for the application of extreme event statistics that usually rely on the assessment of binary time series. In particular, the present work has applied event coincidence analysis (ECA) to quantify simultaneous coincidences between the presence of the different ENSO stages and seasonal extreme, i.e., wet and dry, precipitation events during different seasons following an ENSO event. It has been shown for El Niño, that EP events usually coincide with spatially coherent extreme precipitation signals while the effects of CP events are more dispersed along the entire globe. This observation fits with the assumption put into the definition of the transitivity based discrimination (Chap. 4) that the teleconnective structure of El Niño (and La Niña) differs between the two events. In particular, the most prominent signals, such as reduced rainfall over Indonesia and wet conditions at the North American East Coast, have been recovered and attributed primarily to the EP type of El Niño. For the CP type a variety of previously uncovered coincidences has been observed. This suggests that the observations made in previous studies which did not take any discrimination into account were primarily dominated by the comparably stronger EP signals. Again, in analogy to the procedure in Chap. 4 an assessment of the two proposed La Niña periods has been performed as well and the same qualitative behavior in terms of spatially coherent extreme rainfall responses to the EP type and dispersed patterns for the CP type was observed. This further underlines the proposition that a discrimination of La Niña into two phases is meaningful not only in a statistical sense but also from an impact-based point of view. Future work should apply the concepts used in this work to also study ENSO related extreme impacts on further climatic observables, such as temperature, or to directly relate ENSO and its distinct phases more thoroughly to immediate climate impacts,

such as changes in agricultural yields or water availability. It is further suggested to perform a thorough intercomparison between results obtained from linear and event statistics in order to assess which regions are most affected in terms of extreme climatic responses to the presence of either type and phase of ENSO. As above, it could again be useful to apply the presented framework to future climate projections in order to assess possible changes in the spatial extent and frequency of ENSO related extreme events.

From the assessment of single-layer networks as presented in Chap. 4, Chap. 6 moved towards the study of coupled climate networks that are comprised of more than just a single climatic field. As the dynamics within different parts of the Earth's climatic subsystems, e.g., ocean and atmosphere, are closely entangled, coupled climate networks have been expected to specifically resolve spatial and network topological signatures of these corresponding mutual interactions. Complementing previous studies on qquatorial and Southern hemispheric ocean-atmosphere interactions, the present work has specifically focused on the Northern hemispheric correlation structure between sea surface temperature variability and different layers of geopotential height fields at various altitudes ranging from near-surface to the lower stratosphere. From a combined assessment of the n.s.i. cross-degree density and the n.s.i. local cross-clustering coefficient it was revealed that the ocean-atmosphere interactions display a hierarchical structure in a sense that single spatially confined areas of the ocean correlate with multiple mutually disconnected parts of the atmosphere. Thus, climate network theory proved useful to discriminate the different spatial interaction structures in the two climatic subsystems. The observed hierarchy has then been related to dominant atmospheric patterns, such as the Pacific North American pattern, forcing the ocean on the time scales investigated in this study. After a thorough assessment of local per-node network characteristics, the overall interaction structure between ocean and atmosphere was investigated by an assessment of global interacting network characteristics computed for pairs of the oceanic layer and each of the 18 layers of geopotential height. It has been shown that such global network characteristics discriminate well different parts of the atmosphere with respect to their interaction with the ocean. A multitude of suggestions for future research has been proposed in Sec. 6.6 of the respective chapter. First, it was noted that indirect (partial) correlations or common driver effects within each of the fields were not accounted for when constructing the coupled climate networks. Such possible influences could in the future be addressed by, conditioning out these possible influences by means of information theoretic approaches [Runge et al., 2012a, 2013] and investigating the resulting causal effect networks [Kretschmer et al., 2016, Runge et al., 2015]. Another possibility to rule out the above mentioned effects could be to analyze ensemble simulations of general circulation models and, thus, directly assess the likelihood of an influence of certain possible common drivers on the observables

under study. In order to further investigate the influence of spatio-temporal autocorrelation on the outcome of the present analysis the network could alternatively be constructed by estimating pairwise thresholds from surrogate data as proposed by Paluš et al. [2011]. Ultimately, it remains to note that the analysis presented in this work was only performed for the pairwise correlation between one atmospheric layer and the ocean. Future work could investigate possibilities to refine the coupled climate networks framework in order to also quantify interactions between more than two subnetworks. For the area of study covered in this work, it might for example be of special interest to also consider Arctic sea ice as an additional observable in the network construction and to study possible changes in correlations between ocean and atmosphere over the course of the past decades that may have been induced by processes related to the Arctic amplification.

From the observations drawn in Pt. I and the corresponding acknowledgment that most networks found in nature or the real-world are in fact spatially embedded, the more theory-focused Part II of this work opened in Chap. 7 with a further methodological development that specifically addresses these effects of spatial embedding on the topological quantification and corresponding classification of an arbitrary set of such networks. Specifically, a hierarchy of null models has been proposed that generate (spatially) constrained surrogates of a given network which allow to quantify whether and to what extent topological characteristics of a spatially embedded network are already predetermined by its embedding in some metric space. Specifically, two models closely modeled after the commonly studied Erdős-Rényi random graph and the Configuration Model have been introduced that generate surrogates by iteratively rewiring the original networks such that the average degree or degree distribution is preserved. On top of that, a novel surrogate construction scheme denoted as GeoModelI additionally preserves the link length distribution while a second scheme, GeoModelII, also preserves the same property for each node individually. A study of six real-world and two synthetic networks revealed two different classes of spatially embedded networks, i.e., one where spatial embedding largely determines global network characteristics and one where their respective numerical values are largely independent of the nodes embedding in a metric space. Specifically, it has been speculated that, based on the observations in this study, macroscopic characteristics of networks with a high degree of planarity display an increased tendency to be affected by spatial embedding. However, a thorough and further testing of this hypothesis was suggested as a subject of future research. For this purpose, it might be useful to extend GeoModelII by imposing additional constraints on the rewiring process such that the degree of planarity of the surrogates corresponds to that of the original network. A second, more conceptual, issue that should be addressed in future research is that of the estimation of the tolerance parameter which quantifies allowed length deviations between old and newly rewired links in the generation of

the surrogate networks. Up to now the tolerance was estimated numerically by an optimization such that the Hamming distance between the original and the surrogate networks is as large as possible while the corresponding cumulative link length distributions are still statistically indistinguishable. Future work should investigate in more detail the interplay between the tolerance and, e.g., the shape of the cumulative distribution function of link lengths as well as the number of nodes and links in order to possibly derive an analytical estimation based on the above quantities.

In the second half of Pt. II, Chap. 8 ultimately introduced a novel metric to discriminate different types or classes of networks that is based on an information-theoretic notion of statistical complexity. In particular, the concept of a complexity-entropy plane has been transferred from its application to distinguish different types of chaotic and stochastic processes in time series analysis to the case of complex networks. The corresponding entropic measure has been defined based on random walks through the network, a notion that is closely entangled with that of network navigability which again has often been related to a network's complexity. From this definition, the statistical complexity measure has then been based on a product of that entropy as a measure of information content and the Jenson-Shannon divergence between the network under study and an appropriately randomized reference state as a measure of disequilibrium. An application to 29 different real-world networks with classes assigned by context showed that certain classes of network are located in distinct areas of the resulting complexity-entropy plane. Specifically, it has been revealed that connectome networks display among the highest statistical complexity while, e.g., infrastructure networks show significantly lower values. Generally, the proposed metric appears to serve as a proper discriminator between different classes and related functions of complex networks and is due to its generic nature universally applicable to a large variety of systems. It has been proposed that future work should investigate in more detail the discriminating power of the statistical complexity for a broader set of real world complex networks by, e.g., a further generalization to the case of weighted and/or directed networks. For this purpose, more emphasis must then be put into the definition of the reference networks, which were for now assumed to be just a randomized correspondent of each specific network under study. Linking back to Pt. II of this work, Chap. 8 further presented the possibility to utilize the proposed framework to appropriately estimate thresholds for the construction of functional climate networks such that the statistical network complexity is maximized. This procedure should be taken as a complement to existing approaches of estimating thresholds, such as the most commonly employed one of choosing a desired network link density.

In summary, the observations drawn from the functional climate network analyses in Pt. I show that complex network theory provides valuable complements to

standard tools of classical linear statistics. While in the present case lower-order network characteristics, i.e., the degree, display high similarities with classical results from statistical climatology, network measures of higher order, e.g., transitivity or clustering coefficient, unravel previously hidden structures in the climate system. Moreover, being not restricted to the study of climatic data sets, the metrics and frameworks presented in this work should also prove useful for the purpose of data analysis in many other disciplines, such as neuroscience or economics.

Additionally, the two frameworks that were proposed in Pt. II of this work, i.e, the models to construct surrogates of spatially embedded networks and the novel metric of statistical network complexity, are relevant to the study of complex networks across many disciplines as well. First, since many real-world complex networks are in fact spatial networks, the approaches proposed here help to disentangle the underlying complex system structure from the spatial embedding of nodes in many types of systems, ranging from infrastructure, over communication, to economic networks. Second, the statistical complexity entropy-plane for networks expands the understanding of complex topological structures and helps to quantify varying degrees of complexity in various systems. As already partly illustrated in this work, the presented approach should again be useful for many fields of (applied) complex network science, such as neuro-, social or (as shown here) even climate science.

Appendix A.

A.1. Coincidences between ENSO periods and extreme rainfall for different quantiles

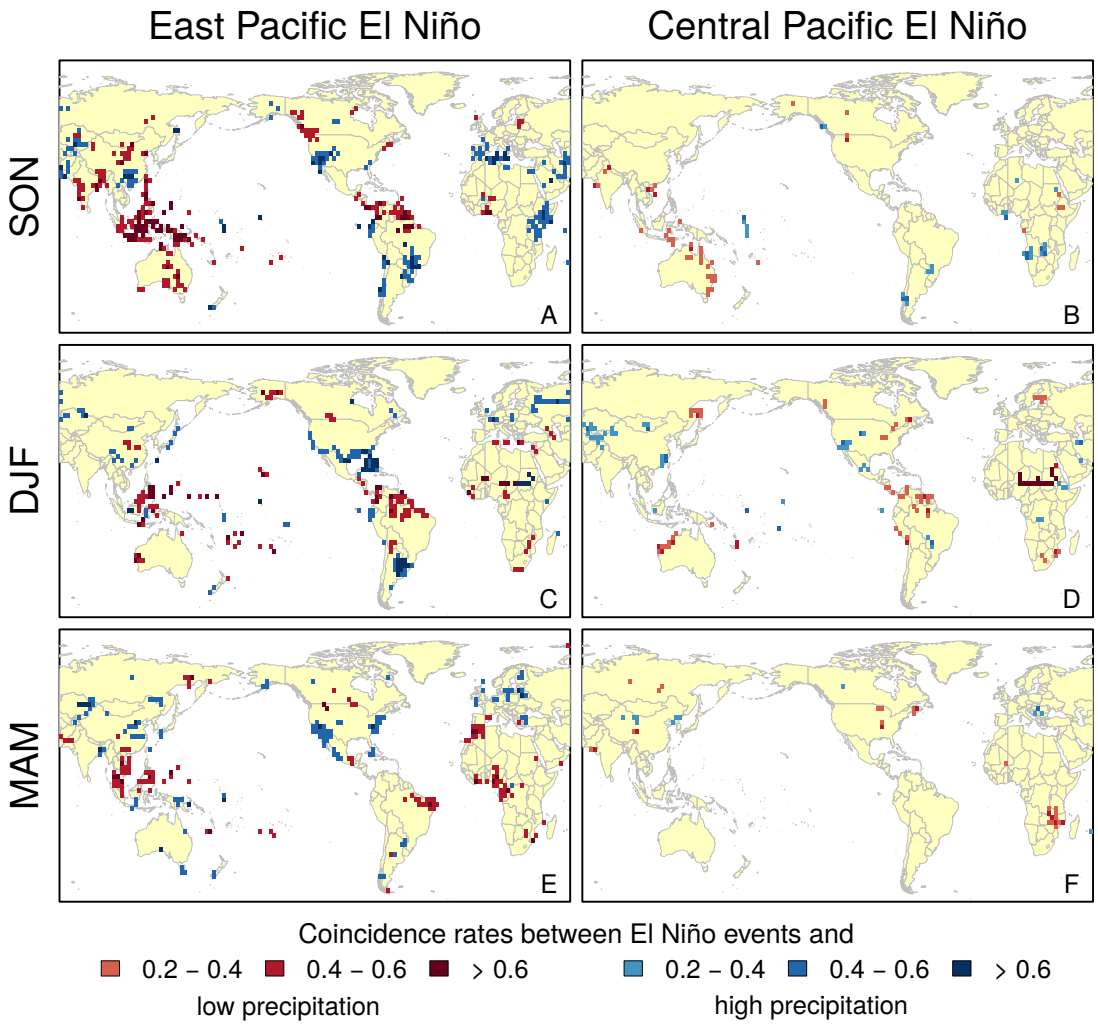

Figure A.1.: The same as Fig. 5.3 but for seasonal precipitation sums below (above) the 15th (85th) percentile of all years from 1951 to 2014.

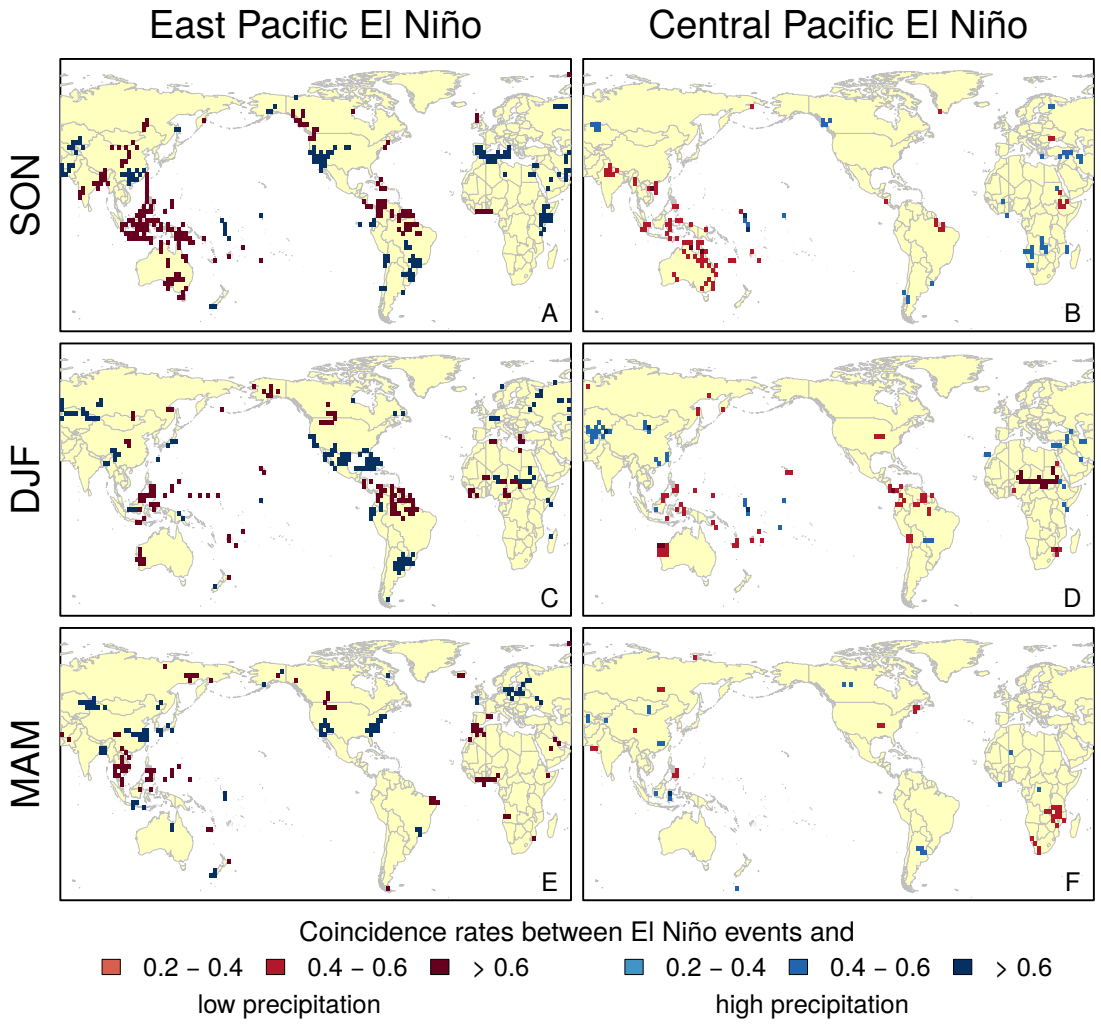

Figure A.2.: The same as Fig. 5.3 but for seasonal precipitation sums below (above) the 25th (75th) percentile of all years from 1951 to 2014.

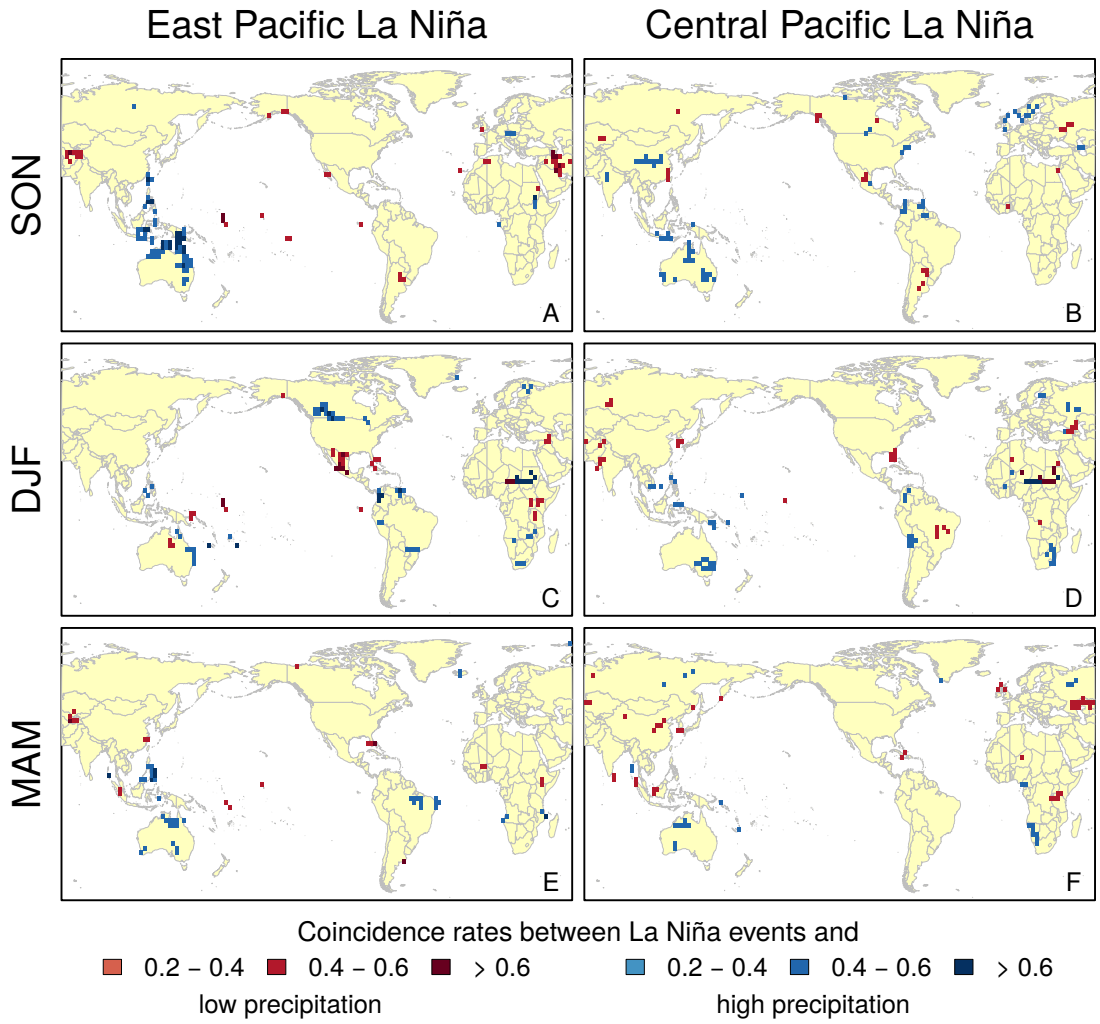

Figure A.3.: The same as Fig. 5.4 but for seasonal precipitation sums below (above) the 15th (85th) percentile of all years from 1951 to 2014.

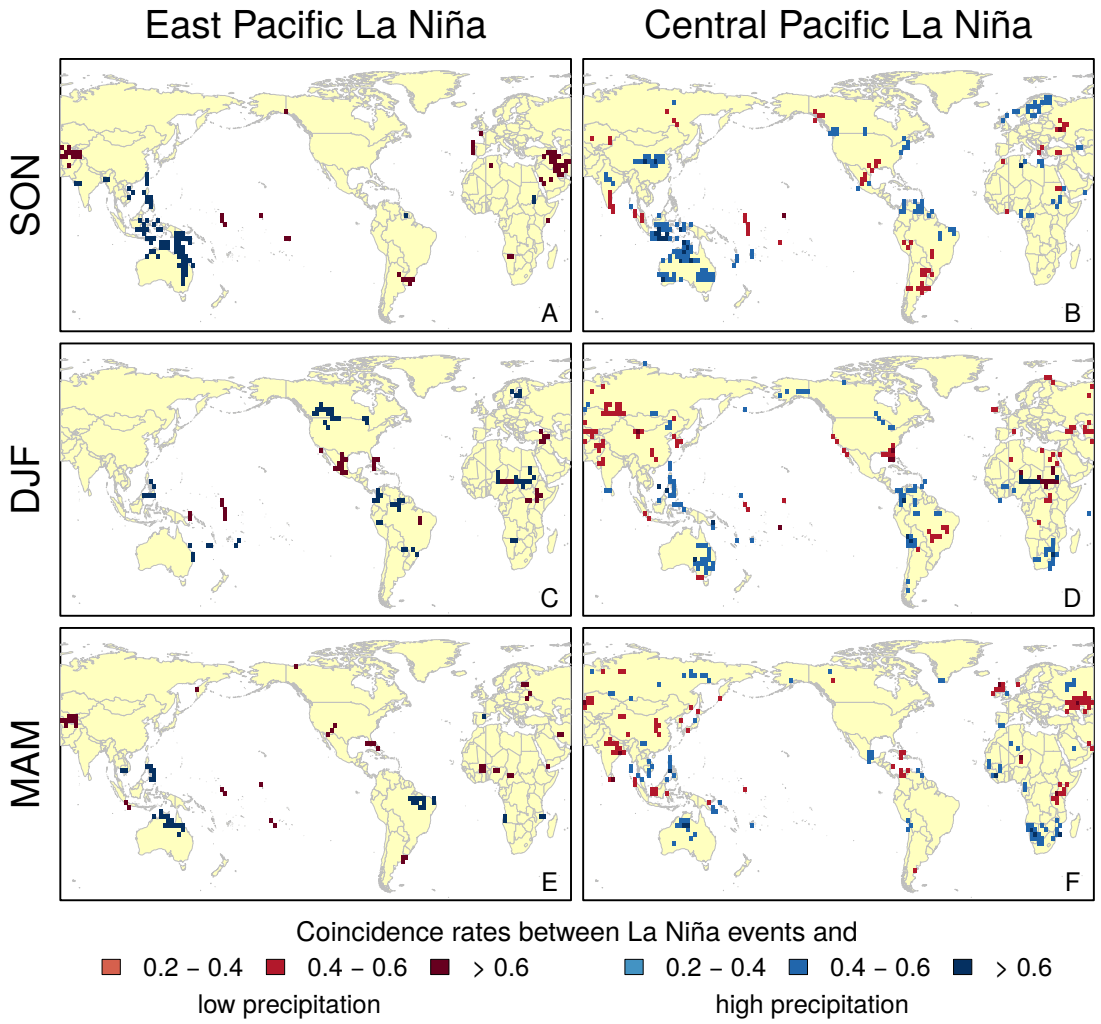

Figure A.4.: The same as Fig. 5.4 but for seasonal precipitation sums below (above) the 25th (75th) percentile of all years from 1951 to 2014.

A.2. Similarities between leading coupled patterns and n.s.i. cross-degree densities

Following Donges et al. [2015b], the relationships between degree-related weighted and unweighted measures for coupled climate networks and the corresponding coupled patterns from weighted MCA are derived. In this work, coupled patterns $\vec{p}_s^{\,(m)}$ and $\vec{p}_h^{\,(m)}$ (see Eqs. (6.3) and (6.4)) are computed from the weighted cross-correlation matrix (Eq. (6.2)) between two climatological fields with elements

$$\rho^w_{s_n h_l} = w_{s_n h_l} \rho_{s_n h_l} = \sqrt{\cos \lambda_{s_n} \cos \lambda_{h_l}} \rho_{s_n h_m} \tag{A.1}$$

with $w_{s_n h_l} = \sqrt{\cos \lambda_{s_n} \cos \lambda_{h_l}}$ according to Eq. (6.2). The weighted cross-correlation matrix can be expanded in terms of coupled patterns and singular values as

$$\rho^w_{s_n h_l} = \sum_{m=1}^{\mathcal{R}} \sigma_k p^{(m)}_{s_n} p^{(m)}_{h_l}. \tag{A.2}$$

This implies that the unweighted cross-correlation matrix can be expressed as

$$\rho_{s_n h_l} = \frac{1}{w_{s_n h_l}} \sum_{m=1}^{\mathcal{R}} \sigma_k p^{(m)}_{s_n} p^{(m)}_{h_l}. \tag{A.3}$$

Since all coupled climate network measures are based on $\rho_{s_n h_l}$, degree-based measures can be written as functions of singular values and coupled patterns from weighted MCA as well. For the unweighted cross-degree, one obtains

$$k^s_i = \sum_{n=1}^{N_{xs}} \Theta\left(|\rho_{s_n h_l}| - T_{sh}\right) = \sum_{n=1}^{N_{xs}} \Theta\left(\frac{1}{w_{s_n h_l}} \left| \sum_{m=1}^{\mathcal{R}} \sigma_k p^{(m)}_{s_n} p^{(m)}_{h_l} \right| - T_{sh}\right) \tag{A.4}$$

and analogously

$$k^h_i = \sum_{l=1}^{N_{xh}} \Theta\left(|\rho_{s_n h_l}| - T_{sh}\right) = \sum_{l=1}^{N_{xh}} \Theta\left(\frac{1}{w_{s_n h_l}} \left| \sum_{m=1}^{\mathcal{R}} \sigma_k p^{(m)}_{s_n} p^{(m)}_{h_l} \right| - T_{sh}\right). \tag{A.5}$$

Similarly, for the n.s.i. cross-degree it follwos that

$$k^{s*}_i = \sum_{n=1}^{N_{xs}} w_n \Theta\left(|\rho_{s_n h_l}| - T_{sh}\right) = \sum_{n=1}^{N_{xs}} w_n \Theta\left(\frac{1}{w_{s_n h_l}} \left| \sum_{m=1}^{\mathcal{R}} \sigma_k p^{(m)}_{s_n} p^{(m)}_{h_l} \right| - T_{sh}\right) \tag{A.6}$$

and

$$k_i^{h*} = \sum_{l=1}^{N_{x_h}} w_l \Theta\left(|\rho_{s_n h_l}| - T_{sh}\right) = \sum_{l=1}^{N_{x_h}} w_l \Theta\left(\frac{1}{w_{s_n h_l}} \left| \sum_{m=1}^{\mathcal{R}} \sigma_k p_{s_n}^{(m)} p_{h_l}^{(m)} \right| - T_{sh}\right). \quad \text{(A.7)}$$

Analogous results hold for the cross-degree and n.s.i. cross-degree densities, respectively.

In the following, the focus lies on the n.s.i. cross-degree k_i^{s*} for illustration, while all results hold for the other degree-related measures as well. If the leading pair of coupled patterns explains a large fraction of the cross-covariance between both climatological fields with $\sigma_1 \gg \sigma_2$ (as for the climatological fields investigated in this study), one can approximate

$$k_i^{s*} \approx \sum_{n=1}^{N_{x_s}} w_n \Theta\left(\frac{\sigma_1}{w_{s_n h_l}} \left| p_{s_n}^{(1)} p_{h_l}^{(1)} \right| - T_{sh}\right) \quad \text{(A.8)}$$

Elements of the leading coupled pattern $p_{h_l}^{(1)}$ contribute to this sum if $|p_{h_l}^{(1)}| \geq T_{sh} w_{s_n h_l} / \sigma_1 |p_{s_n}^{(1)}|$. Hence, a larger $|p_{h_l}^{(1)}|$ increases the odds for a larger k_i^{s*} to arise, implying a positive correlation between the absolute coefficient of the leading coupled pattern $|p_{h_l}^{(1)}|$ and the n.s.i. cross-degree k_i^{s*} as well as the n.s.i. cross-degree density $\kappa_i^{s*} = k_i^{s*} / \mathcal{W}_s$, as it is observed in this study.

A.3. Sensitivity of the power law relationship

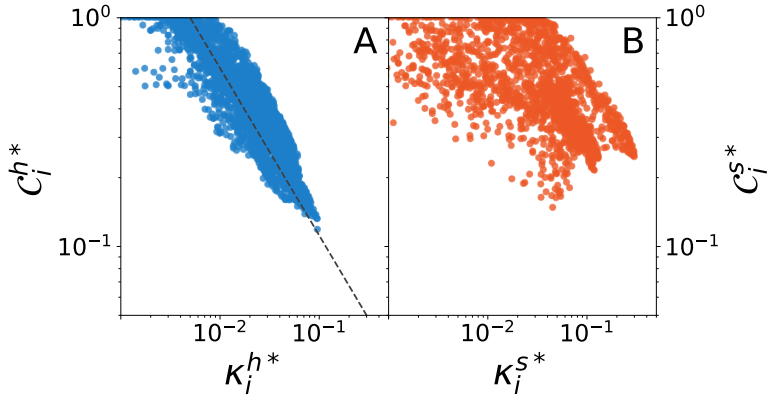

Figure A.5.: The same as Fig. 6.6 for a network with an internal link density $d_s = d_h = 0.02$ and a cross-link density $d_{sh} = 0.01$. The dashed line in (A) indicates again the relationship $\mathcal{C}_i^{h*}(\kappa_i^{h*}) \sim (\kappa_i^{h*})^{-\alpha}$ (here with $\alpha = 0.90$) expected for traditional network measures $C_i(k_i)$ in the case of hierarchical network structures [Ravasz et al., 2002, Ravasz and Barabási, 2003].

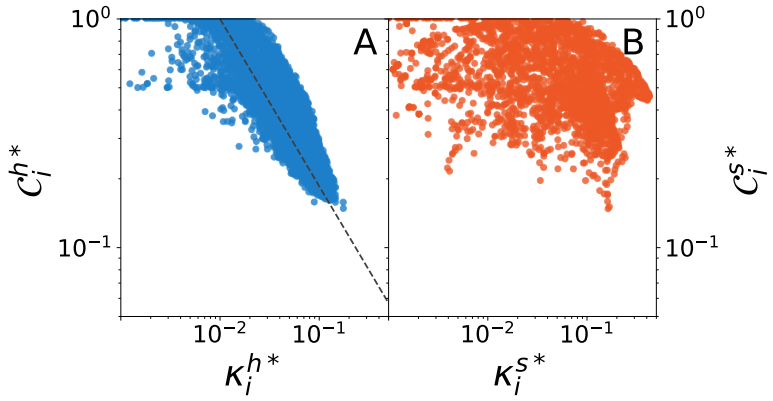

Figure A.6.: The same as Fig. A.5 for a network with an internal link density $d_s = d_h = 0.05$ and a cross-link density $d_{sh} = 0.025$. The power-law exponent for this case reads $\alpha = 0.83$.

The relationship between κ_i^{h*} (κ_i^{s*}) and $\mathcal{C}_i^{h*}$ ($\mathcal{C}_i^{s*}$) is assessed as analogue to Fig. 6.6 but for different choices of internal $d_s = d_h$ and cross-link densities d_{sh}, respectively. Fig. A.5 shows the results for $d_s = d_h = 0.02$ and $d_{sh} = 0.01$. Fig. A.6 displays the same relationship for even larger link densities $d_s = d_h = 0.05$ and $d_{sh} = 0.025$. In both cases one observes a similar power-law shaped relationship $\mathcal{C}_i^{h*}(\kappa_i^{h*}) \sim (\kappa_i^{h*})^{-\alpha}$. Since the results presented in Fig. A.5 and Fig. A.6 as well as Fig. 6.6 proof to be

qualitatively very similar, this part of the analysis is considered to be sufficiently robust against the actual choice of internal and cross-link densities.

A.4. Hierarchical structures in the 200 mbar HGT field

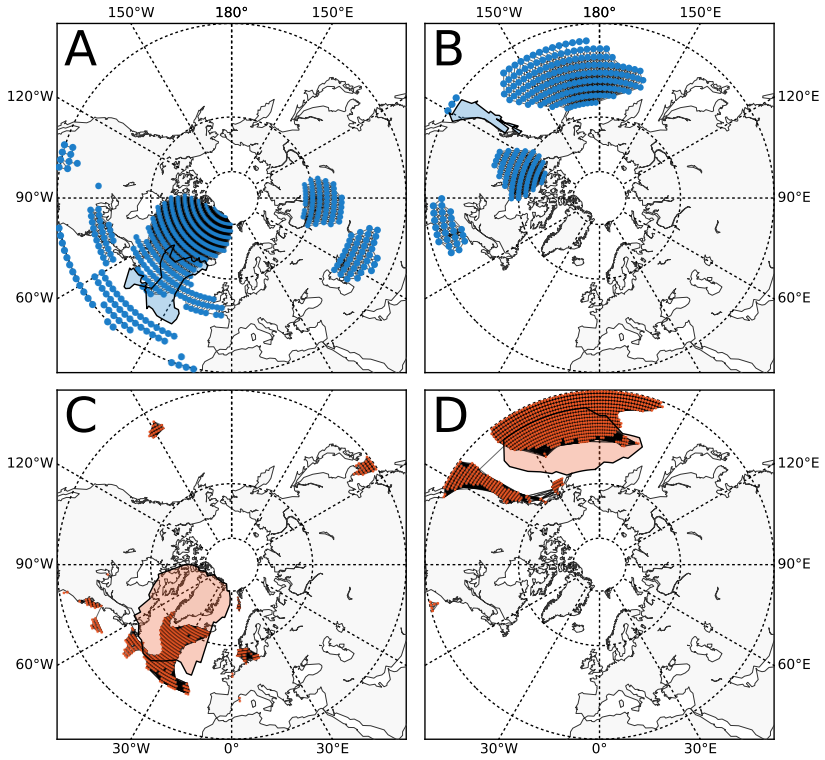

Figure A.7.: Visualization of a selection of nodes that are relevant for the observed hierarchical network structure. (A,B) Two clusters of nodes in the SST field (blue and orange shaded polygons) that show positive values of κ_i^{h*} with the 200 mbar HGT field (compare Fig. 6.3E). Correspondingly, coloured scatter points denote nodes in the HGT field that the considered SST nodes are connected with. All links that mutually connect the resulting HGT nodes are displayed as well. (C,D) The same for the two largest patches of nodes in the 200 mbar HGT field that where detected in (A,B). Coloured scatter points now indicate all nodes in the SST field that are connected with these patches.

A.5. References to the networks that are used to study statistical complexity

Name	Reference
Sheep	Hass [1991]
Rhesus	Sade [2004]
Kangaroo	Grant [1973]
Mac	Fedigan and Asquith [1991]
Bison	Lott [1979]
Zebra n	Sundaresan et al. [2007]
Cattle	Schein and Fohrman [1955]
Dolphins	Lusseau et al. [2003]
Autobahn	Kaiser and Hilgetag [2004]
USairport500	Colizza et al. [2007]
USairport 2010	Opsahl [2011]
Openflights	Opsahl [2011]
Rome99	`dis.uniroma1.it/challenge9`
South-Africa	Hobson [1906]
American Revolution	`konect.uni-koblenz.de/networks`
Club-Membership	Faust [1997]
Corporate-Leadership	Barnes and Burkett [2010]
Rhesus Brain 1	Harriger et al. [2012]
Rhesus Brain 2	Markov et al. [2014b]
Mouse Retina 1	Helmstaedter et al. [2013]
Mixed Species Brain 1	Reus and Heuvel [2013]
Rhesus Cerebral Cortex 1	Markov et al. [2014a]
C Elegans Neural Male 1	Jarrell et al. [2012]6
Rattus Norvegicus Brain 3	Bota and Swanson [2007]
Rhesus Interarea Cortical 2	Markov et al. [2013]
Rattus Norvegicus Brain 2	Bota and Swanson [2007]
Rattus Norvegicus Brain 1	Bota and Swanson [2007]
Mouse Brain 1	Oh et al. [2014]
C Elegans Herm Pharynx 1	Varshney et al. [2011]

Table A.1.: References to the networks that are used to study statistical complexity

Bibliography

O. Aladenola, A. Cashman, and D. Brown. Impact of El Niño and Climate Change on Rainwater Harvesting in a Caribbean State. *Water Resources Management*, 30 (10):3459–3473, 2016. doi: 10.1007/s11269-016-1362-2.

R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97, 2002. doi: 10.1103/RevModPhys.74.47.

R. Albert, I. Albert, and G. L. Nakarado. Structural vulnerability of the North American power grid. *Physical Review E*, 69(2):025103, 2004. doi: 10.1103/PhysRevE.69.025103.

R. Allan, J. Lindesay, and D. Parker. El Niño Southern Oscillation and climatic variability. *Oceanographic Literature Review*, 6(44):555, 1997.

R. P. Allan and B. J. Soden. Atmospheric Warming and the Amplification of Precipitation Extremes. *Science*, 321(5895):1481–1484, 2008. doi: 10.1126/science.1160787.

L. a. N. Amaral and J. M. Ottino. Complex networks. *The European Physical Journal B - Condensed Matter and Complex Systems*, 38(2):147–162, 2004. doi: 10.1140/epjb/e2004-00110-5.

L. a. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley. Classes of small-world networks. *Proceedings of the National Academy of Sciences*, 97(21):11149–11152, 2000. doi: 10.1073/pnas.200327197.

S.-I. An and B. Wang. The Forced and Intrinsic Low-Frequency Modes in the North Pacific. *Journal of Climate*, 18(6):876–885, 2005. doi: 10.1175/JCLI-3298.1.

K. Anand and G. Bianconi. Entropy measures for networks: Toward an information theory of complex topologies. *Physical Review E*, 80(4):045102, 2009. doi: 10.1103/PhysRevE.80.045102.

P. W. Anderson. Is complexity physics? Is it science? What is it? *Physics Today*, 44(7):9–11, 1991.

G. Ansmann and K. Lehnertz. Constrained randomization of weighted networks. *Physical Review E*, 84(2):026103, 2011. doi: 10.1103/PhysRevE.84.026103.

J. Arblaster, G. Meehl, and A. Moore. Interdecadal modulation of Australian rainfall. *Climate Dynamics*, 18(6):519–531, 2002. doi: 10.1007/s00382-001-0191-y.

Y. Artzy-Randrup and L. Stone. Generating uniformly distributed random networks. *Physical Review E*, 72(5):056708, 2005. doi: 10.1103/PhysRevE.72.056708.

K. Ashok and T. Yamagata. Climate change: The El Niño with a difference. *Nature*, 461(7263):481–484, 2009. doi: 10.1038/461481a.

K. Ashok, S. K. Behera, S. A. Rao, H. Weng, and T. Yamagata. El Niño Modoki and its possible teleconnection. *Journal of Geophysical Research: Oceans*, 112(C11): C11007, 2007. doi: 10.1029/2006JC003798.

G. Bagler. Analysis of the airport network of India as a complex weighted network. *Physica A: Statistical Mechanics and its Applications*, 387(12):2972–2980, 2008. doi: 10.1016/j.physa.2008.01.077.

J. R. Banavar, A. Maritan, and A. Rinaldo. Size and form in efficient transportation networks. *Nature*, 399(6732):130–132, 1999. doi: 10.1038/20144.

A.-L. Barabasi. Scale-Free Networks: A Decade and Beyond. *Science*, 325(5939): 412–413, 2009. doi: 10.1126/science.1173299.

A.-L. Barabási and R. Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, 1999. doi: 10.1126/science.286.5439.509.

W. Barfuss, J. F. Donges, M. Wiedermann, and W. Lucht. Sustainable use of renewable resources in a stylized social–ecological network model under heterogeneous resource distribution. *Earth System Dynamics*, 8(2):255–264, 2017. doi: 10.5194/esd-8-255-2017.

R. Barnes and T. Burkett. Structural redundancy and multiplicity in corporate networks. *Connections*, 30(2):4–20, 2010.

L. Barnett, E. Di Paolo, and S. Bullock. Spatially embedded random networks. *Physical Review E*, 76(5):056115, 2007. doi: 10.1103/PhysRevE.76.056115.

A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences*, 101(11):3747–3752, 2004a. doi: 10.1073/pnas.0400087101.

A. Barrat, M. Barthélemy, and A. Vespignani. Weighted Evolving Networks: Coupling Topology and Weight Dynamics. *Physical Review Letters*, 92(22):228701, 2004b. doi: 10.1103/PhysRevLett.92.228701.

M. Barreiro, A. C. Marti, and C. Masoller. Inferring long memory processes in the climate network via ordinal pattern analysis. *Chaos*, 21(1):013101, 2011.

M. Barthélemy. Spatial networks. *Physics Reports*, 499(1–3):1–101, 2011. doi: 10.1016/j.physrep.2010.11.002.

S. Bialonski, M.-T. Horstmann, and K. Lehnertz. From brain to earth and climate systems: Small-world interaction networks or not? *Chaos*, 20(1):013134, 2010. doi: 10.1063/1.3360561.

G. Bianconi. The entropy of randomized network ensembles. *Europhysics Letters*, 81(2):28005, 2008. doi: 10.1209/0295-5075/81/28005.

J. Bjerknes. Atmospheric teleconnections from the equatorial pacific. *Monthly Weather Review*, 97(3):163–172, 1969. doi: 10.1175/1520-0493(1969)097<0163:ATFTEP>2.3.CO;2.

S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424(4-5):175–308, 2006. doi: 10.1016/j.physrep.2005.10.009.

S. Boccaletti, G. Bianconi, R. Criado, C. I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin. The structure and dynamics of multilayer networks. *Physics Reports*, 544(1):1–122, 2014. doi: 10.1016/j.physrep.2014.07.001.

N. Boers, B. Bookhagen, H. M. J. Barbosa, N. Marwan, J. Kurths, and J. A. Marengo. Prediction of extreme floods in the eastern Central Andes based on a complex networks approach. *Nature Communications*, 5(5199), 2014a. doi: 10.1038/ncomms6199.

N. Boers, B. Bookhagen, N. Marwan, J. Kurths, and J. Marengo. Complex networks identify spatial patterns of extreme rainfall events of the South American Monsoon System. *Geophysical Research Letters*, 40(16):4386–4392, 2013. doi: 10.1002/grl.50681.

N. Boers, A. Rheinwalt, B. Bookhagen, H. M. J. Barbosa, N. Marwan, J. Marengo, and J. Kurths. The South American rainfall dipole: A complex network analysis of extreme events. *Geophysical Research Letters*, 41(20):7397–7405, 2014b. doi: 10.1002/2014GL061829.

N. Boers, A. Rheinwalt, B. Bookhagen, N. Marwan, and J. Kurths. A Complex Network Approach to Investigate the Spatiotemporal Co-variability of Extreme Rainfall. In *Machine Learning and Data Mining Approaches to Climate Science*,

pages 163–174. Springer International Publishing, 2015. doi: 10.1007/978-3-319-17220-0_15.

D. Bonchev. *Information theoretic indices for characterization of chemical structures*. Number 5. Research Studies Press, 1983.

D. Bonchev and G. A. Buck. Quantitative measures of network complexity. In *Complexity in chemistry, biology, and ecology*, pages 191–235. Springer, 2005.

B. Bookhagen and M. R. Strecker. Spatiotemporal trends in erosion rates across a pronounced rainfall gradient: Examples from the southern Central Andes. *Earth and Planetary Science Letters*, 327–328:97–110, 2012. doi: 10.1016/j.epsl.2012.02.005.

M. Bota and L. W. Swanson. Online workbenches for neural network connections. *The Journal of Comparative Neurology*, 500(5):807–814, 2007. doi: 10.1002/cne.21209.

C. S. Bretherton, C. Smith, and J. M. Wallace. An Intercomparison of Methods for Finding Coupled Patterns in Climate Data. *Journal of Climate*, 5(6):541–560, 1992. doi: 10.1175/1520-0442(1992)005<0541:AIOMFF>2.0.CO;2.

C. E. Buell. Correlation Functions for Wind and Geopotential on Isobaric Surfaces. *Journal of Applied Meteorology*, 11(1):51–59, 1972. doi: 10.1175/1520-0450(1972)011<0051:CFFWAG>2.0.CO;2.

S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin. Catastrophic cascade of failures in interdependent networks. *Nature*, 464(7291):1025–1028, 2010. doi: 10.1038/nature08932.

E. Bullmore and O. Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, 2009. doi: 10.1038/nrn2575.

L. Buzna, L. Issacharoff, and D. Helbing. The evolution of the topology of high-voltage electricity networks. *International Journal of Critical Infrastructures*, 5 (1-2):72–85, 2009. doi: 10.1504/IJCIS.2009.02285.

W. Cai and T. Cowan. La Niña Modoki impacts Australia autumn rainfall variability. *Geophysical Research Letters*, 36(12):L12805, 2009. doi: 10.1029/2009GL037885.

W. Cai, S. Borlace, M. Lengaigne, P. van Rensch, M. Collins, G. Vecchi, A. Timmermann, A. Santoso, M. J. McPhaden, L. Wu, M. H. England, G. Wang, E. Guilyardi, and F.-F. Jin. Increasing frequency of extreme El Niño events due to greenhouse warming. *Nature Climate Change*, 4(2):111–116, 2014. doi: 10.1038/nclimate2100.

X. Calbet and R. López-Ruiz. Tendency towards maximum complexity in a nonequilibrium isolated system. *Physical Review E*, 63(6):066116, 2001. doi: 10.1103/PhysRevE.63.066116.

P. Camberlin, S. Janicot, and I. Poccard. Seasonality and atmospheric dynamics of the teleconnection between African rainfall and tropical sea-surface temperature: Atlantic vs. ENSO. *International Journal of Climatology*, 21(8):973–1005, 2001. doi: 10.1002/joc.673.

R. F. i. Cancho and R. V. Solé. Optimization in Complex Networks. In R. Pastor-Satorras, M. Rubi, and A. Diaz-Guilera, editors, *Statistical Mechanics of Complex Networks*, number 625 in Lecture Notes in Physics, pages 114–126. Springer Berlin Heidelberg, 2003. ISBN 978-3-540-40372-2 978-3-540-44943-0. doi: 10.1007/978-3-540-44943-0_7.

A. Capocci, V. D. P. Servedio, F. Colaiori, L. S. Buriol, D. Donato, S. Leonardi, and G. Caldarelli. Preferential attachment in the growth of social networks: The internet encyclopedia Wikipedia. *Physical Review E*, 74(3):036116, 2006. doi: 10.1103/PhysRevE.74.036116.

E. Castillo. Functional Networks. *Neural Processing Letters*, 7(3):151–159, 1998. doi: 10.1023/A:1009656525752.

S. H. Y. Chan, R. V. Donner, and S. Lämmer. Urban road networks — spatial networks with universal geometric features? *The European Physical Journal B - Condensed Matter and Complex Systems*, 84(4):563–577, 2011. doi: 10.1140/epjb/e2011-10889-3.

Y. Chen, S. Wang, C. C. Hilgetag, and C. Zhou. Trade-off between Multiple Constraints Enables Simultaneous Formation of Modules and Hubs in Neural Systems. *PLoS Computational Biology*, 9(3):e1002937, 2013. doi: 10.1371/journal.pcbi.1002937.

F. H. S. Chiew, T. C. Piechota, J. A. Dracup, and T. A. McMahon. El Niño/Southern Oscillation and Australian rainfall, streamflow and drought: Links and potential for forecasting. *Journal of Hydrology*, 204(1–4):138–149, 1998. doi: 10.1016/S0022-1694(97)00121-2.

V. Colizza, R. Pastor-Satorras, and A. Vespignani. Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nature Physics*, 3(4):276–282, 2007. doi: 10.1038/nphys560.

L. d. F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas. Characterization of complex networks: A survey of measurements. *Advances in Physics*, 56(1):167–242, 2007. doi: 10.1080/00018730601170527.

D. P. Croft, R. James, and J. Krause. *Exploring animal social networks*. Princeton University Press, 2008.

P. Crucitti, V. Latora, and M. Marchiori. A topological analysis of the Italian electric power grid. *Physica A: Statistical Mechanics and its Applications*, 338(1–2):92–97, 2004. doi: 10.1016/j.physa.2004.02.029.

P. Crucitti, V. Latora, and S. Porta. Centrality measures in spatial networks of urban streets. *Physical Review E*, 73(3):036125, 2006. doi: 10.1103/PhysRevE.73.036125.

A. Czaja and C. Frankignoul. Influence of the North Atlantic SST on the atmospheric circulation. *Geophysical Research Letters*, 26(19):2969–2972, 1999. doi: 10.1029/1999GL900613.

S. Dafermos and A. Nagurney. Sensitivity analysis for the asymmetric network equilibrium problem. *Mathematical Programming*, 28(2):174–184, 1984. doi: 10.1007/BF02612357.

A. Dai and T. M. L. Wigley. Global patterns of ENSO-induced precipitation. *Geophysical Research Letters*, 27(9):1283–1286, 2000. doi: 10.1029/1999GL011140.

J. Davidsen, H. Ebel, and S. Bornholdt. Emergence of a Small World from Local Interactions: Modeling Acquaintance Networks. *Physical Review Letters*, 88(12):128701, 2002. doi: 10.1103/PhysRevLett.88.128701.

A. De Montis, M. Barth, A. Chessa, A. Vespignani, and others. The structure of interurban traffic: a weighted network analysis. *Environment and Planning B: Planning and Design*, 34(5):905–924, 2007.

M. Dehmer and A. Mowshowitz. A history of graph entropy measures. *Information Sciences*, 181(1):57–78, 2011. doi: 10.1016/j.ins.2010.08.041.

M. Dehmer, N. Barbarini, K. Varmuza, and A. Graber. A large scale analysis of information-theoretic network complexity measures using chemical structures. *PLoS One*, 4(12):e8057, 2009.

M. Dehmer, M. Grabner, and K. Varmuza. Information Indices with High Discriminative Power for Graphs. *PLoS One*, 7(2):e31214, 2012. doi: 10.1371/journal.pone.0031214.

J. I. Deza, M. Barreiro, and C. Masoller. Inferring interdependencies in climate networks constructed at inter-annual, intra-season and longer time scales. *The European Physical Journal Special Topics*, 222(2):511–523, 2013. doi: 10.1140/epjst/e2013-01856-5.

H. F. Diaz, M. P. Hoerling, and J. K. Eischeid. ENSO variability, teleconnections and climate change. *International Journal of Climatology*, 21(15):1845–1862, 2001. doi: 10.1002/joc.631.

M. Dima and G. Lohmann. A Hemispheric Mechanism for the Atlantic Multidecadal Oscillation. *Journal of Climate*, 20(11):2706–2719, 2007. doi: 10.1175/JCLI4174.1.

J. F. Donges, Y. Zou, N. Marwan, and J. Kurths. The backbone of the climate network. *Europhysics Letters*, 87(4):48007, 2009a. doi: 10.1209/0295-5075/87/48007.

J. F. Donges, Y. Zou, N. Marwan, and J. Kurths. Complex networks in climate dynamics. *The European Physical Journal Special Topics*, 174(1):157–179, 2009b. doi: 10.1140/epjst/e2009-01098-2.

J. F. Donges, R. V. Donner, K. Rehfeld, N. Marwan, M. H. Trauth, and J. Kurths. Identification of dynamical transitions in marine palaeoclimate records by recurrence network analysis. *Nonlinear Processes in Geophysics*, 18(5):545–562, 2011a. doi: 10.5194/npg-18-545-2011.

J. F. Donges, H. C. H. Schultz, N. Marwan, Y. Zou, and J. Kurths. Investigating the topology of interacting networks. *The European Physical Journal B - Condensed Matter and Complex Systems*, 84(4):635–651, 2011b. doi: 10.1140/epjb/e2011-10795-8.

J. F. Donges, C.-F. Schleussner, J. F. Siegmund, and R. V. Donner. Event coincidence analysis for quantifying statistical interrelationships between event time series. *The European Physical Journal Special Topics*, 225(3):471–487, 2016. doi: 10.1140/epjst/e2015-50233-y.

J. F. Donges, R. V. Donner, M. H. Trauth, N. Marwan, H.-J. Schellnhuber, and J. Kurths. Nonlinear detection of paleoclimate-variability transitions possibly related to human evolution. *Proceedings of the National Academy of Sciences*, 108 (51):20422–20427, 2011c. doi: 10.1073/pnas.1117052108.

J. F. Donges, J. Heitzig, R. V. Donner, and J. Kurths. Analytical framework for recurrence network analysis of time series. *Physical Review E*, 85(4):046105, 2012. doi: 10.1103/PhysRevE.85.046105.

J. F. Donges, J. Heitzig, B. Beronov, M. Wiedermann, J. Runge, Q. Y. Feng, L. Tupikina, V. Stolbova, R. V. Donner, N. Marwan, H. A. Dijkstra, and J. Kurths. Unified functional network and nonlinear time series analysis for complex systems science: The pyunicorn package. *Chaos*, 25(11):113101, 2015a. doi: 10.1063/1.4934554.

J. F. Donges, I. Petrova, A. Loew, N. Marwan, and J. Kurths. How complex climate networks complement eigen techniques for the statistical analysis of climatological data. *Climate Dynamics*, 45(9-10):2407–2424, 2015b. doi: 10.1007/s00382-015-2479-3.

R. V. Donner, Y. Zou, J. F. Donges, N. Marwan, and J. Kurths. Recurrence networks—a novel paradigm for nonlinear time series analysis. *New Journal of Physics*, 12(3):033025, 2010. doi: 10.1088/1367-2630/12/3/033025.

R. V. Donner, M. Small, J. F. Donges, N. Marwan, Y. Zou, R. Xiang, and J. Kurths. Recurrence-based time series analysis by means of complex network methods. *International Journal of Bifurcation and Chaos*, 21(04):1019–1046, 2011. doi: 10.1142/S0218127411029021.

R. V. Donner, M. Wiedermann, and J. F. Donges. Complex network techniques for climatological data analysis. In C. L. Franzke and T. J. O'Kane, editors, *Nonlinear and Stochastic Climate Dynamics*. Cambridge University Press, 2017.

D. R. Easterling, J. L. Evans, P. Y. Groisman, T. R. Karl, K. E. Kunkel, and P. Ambenje. Observed Variability and Trends in Extreme Climate Events: A Brief Review. *Bulletin of the American Meteorological Society*, 81(3):417–425, 2000. doi: 10.1175/1520-0477(2000)081<0417:OVATIE>2.3.CO;2.

H. Ebel, J. Davidsen, and S. Bornholdt. Dynamics of social networks. *Complexity*, 8(2):24–27, 2002.

I. Ebert-Uphoff and Y. Deng. Causal Discovery for Climate Research Using Graphical Models. *Journal of Climate*, 25(17):5648–5665, 2012. doi: 10.1175/JCLI-D-11-00387.1.

E. T. Epule, C. Peng, L. Lepage, and Z. Chen. The causes, effects and challenges of Sahelian droughts: a critical review. *Regional Environmental Change*, 14(1):145–156, 2014. doi: 10.1007/s10113-013-0473-z.

P. Erdös and A. Rényi. On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17–61, 1960.

D. Eroglu, N. Marwan, S. Prasad, and J. Kurths. Finding recurrence networks' threshold adaptively for a specific time series. *Nonlinear Processes in Geophysics*, 21(6):1085–1092, 2014. doi: 10.5194/npg-21-1085-2014.

D. Eroglu, F. H. McRobie, I. Ozken, T. Stemler, K.-H. Wyrwoll, S. F. M. Breitenbach, N. Marwan, and J. Kurths. See–saw relationship of the Holocene East Asian–Australian summer monsoon. *Nature Communications*, 7:12929, 2016. doi: 10.1038/ncomms12929.

L. Eymard and G. Reverdin. Ocean–Atmosphere Interactions. In A. Monaco and P. Prouzet, editors, *Ocean in the Earth System*, pages 105–144. John Wiley & Sons, Inc., 2014. ISBN 978-1-119-00767-8. doi: 10.1002/9781119007678.ch3.

G. Fagiolo, J. Reyes, and S. Schiavo. On the topological properties of the world trade web: A weighted network analysis. *Physica A: Statistical Mechanics and its Applications*, 387(15):3868–3873, 2008. doi: 10.1016/j.physa.2008.01.050.

K. Faust. Centrality in affiliation networks. *Social networks*, 19(2):157–191, 1997.

L. M. Fedigan and P. J. Asquith. *The Monkeys of Arashiyama: Thirty-five Years of Research in Japan and the West*. SUNY Press, 1991. ISBN 978-0-7914-0552-9. Google-Books-ID: 9zwroOXx4WoC.

A. Feng, Z. Gong, Q. Wang, and G. Feng. Three-dimensional air–sea interactions investigated with bilayer networks. *Theoretical and Applied Climatology*, 109(3-4): 635–643, 2012. doi: 10.1007/s00704-012-0600-7.

Q. Y. Feng and H. Dijkstra. Are North Atlantic multidecadal SST anomalies westward propagating? *Geophysical Research Letters*, 41(2):2013GL058687, 2014. doi: 10.1002/2013GL058687.

Q. Y. Feng, J. P. Viebahn, and H. A. Dijkstra. Deep ocean early warning signals of an Atlantic MOC collapse. *Geophysical Research Letters*, 41(16):6009–6015, 2014. doi: 10.1002/2014GL061019.

S. Fortunato. Community detection in graphs. *Physics Reports*, 486(3–5):75–174, 2010. doi: 10.1016/j.physrep.2009.11.002.

C. Frankignoul and N. Sennéchael. Observed Influence of North Pacific SST Anomalies on the Atmospheric Circulation. *Journal of Climate*, 20(3):592–606, 2007. doi: 10.1175/JCLI4021.1.

C. Frankignoul, G. de Coëtlogon, T. M. Joyce, and S. Dong. Gulf Stream Variability and Ocean–Atmosphere Interactions. *Journal of Physical Oceanography*, 31(12): 3516–3529, 2001. doi: 10.1175/1520-0485(2002)031<3516:GSVAOA>2.0.CO;2.

G. Gastineau and C. Frankignoul. Influence of the North Atlantic SST Variability on the Atmospheric Circulation during the Twentieth Century. *Journal of Climate*, 28(4):1396–1416, 2015. doi: 10.1175/JCLI-D-14-00424.1.

M. T. Gastner and M. E. J. Newman. The spatial structure of networks. *The European Physical Journal B - Condensed Matter and Complex Systems*, 49(2): 247–252, 2006. doi: 10.1140/epjb/e2006-00046-8.

M. Ghil, M. R. Allen, M. D. Dettinger, K. Ide, D. Kondrashov, M. E. Mann, A. W. Robertson, A. Saunders, Y. Tian, F. Varadi, and P. Yiou. Advanced Spectral Methods for Climatic Time Series. *Reviews of Geophysics*, 40(1):1003, 2002. doi: 10.1029/2000RG000092.

F. Giannessi and A. Maugeri, editors. *Variational Inequalities and Network Equilibrium Problems*. Springer US, Boston, MA, 1995. ISBN 978-1-4899-1360-9 978-1-4899-1358-6. doi: 10.1007/978-1-4899-1358-6.

M. H. Glantz, R. W. Katz, N. Nicholls, and others. *Teleconnections linking worldwide climate anomalies*, volume 535. Cambridge University Press Cambridge, 1991.

K.-I. Goh, E. Oh, H. Jeong, B. Kahng, and D. Kim. Classification of scale-free networks. *Proceedings of the National Academy of Sciences*, 99(20):12583–12588, 2002. doi: 10.1073/pnas.202301299.

A. Gozolchiani, K. Yamasaki, O. Gazit, and S. Havlin. Pattern of climate network blinking links follows El Niño events. *Europhysics Letters*, 83(2):28005, 2008. doi: 10.1209/0295-5075/83/28005.

A. Gozolchiani, S. Havlin, and K. Yamasaki. Emergence of El Niño as an Autonomous Component in the Climate Network. *Physical Review Letters*, 107(14):148501, 2011. doi: 10.1103/PhysRevLett.107.148501.

H.-F. Graf and D. Zanchettin. Central Pacific El Niño, the "subtropical bridge," and Eurasian climate. *Journal of Geophysical Research: Atmospheres*, 117(D1), 2012. doi: 10.1029/2011JD016493.

T. R. Grant. Dominance and association among members of a captive and a free-ranging group of grey kangaroos (Macropus giganteus). *Animal Behaviour*, 21(3): 449–456, 1973. doi: 10.1016/S0003-3472(73)80004-1.

R. Guimerà, S. Mossa, A. Turtschi, and L. a. N. Amaral. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences*, 102(22):7794–7799, 2005. doi: 10.1073/pnas.0407994102.

R. H. Guthrie, S. J. Mitchell, N. Lanquaye-Opoku, and S. G. Evans. Extreme weather and landslide initiation in coastal British Columbia. *Quarterly Journal of Engineering Geology and Hydrogeology*, 43(4):417–428, 2010. doi: 10.1144/1470-9236/08-119.

B. R. Guzman Herrador, B. F. de Blasio, E. MacDonald, G. Nichols, B. Sudre, L. Vold, J. C. Semenza, and K. Nygård. Analytical studies assessing the

association between extreme precipitation or temperature and drinking water-related waterborne infections: a review. *Environmental Health*, 14:29, 2015. doi: 10.1186/s12940-015-0014-y.

R. Hamming. Error detecting and error correcting codes. *Bell System Technical Journal, The*, 29(2):147–160, 1950. doi: 10.1002/j.1538-7305.1950.tb00463.x.

D. Handorf and K. Dethloff. How well do state-of-the-art atmosphere-ocean general circulation models reproduce atmospheric teleconnection patterns? *Tellus A*, 64, 2012. doi: 10.3402/tellusa.v64i0.19777.

D. Handorf, R. Jaiser, K. Dethloff, A. Rinke, and J. Cohen. Impacts of Arctic sea ice and continental snow cover changes on atmospheric winter teleconnections. *Geophysical Research Letters*, 42(7):2367–2377, 2015. doi: 10.1002/2015GL063203.

A. Hannachi, I. T. Jolliffe, and D. B. Stephenson. Empirical orthogonal functions and related techniques in atmospheric science: A review. *International Journal of Climatology*, 27(9):1119–1152, 2007. doi: 10.1002/joc.1499.

L. Harriger, M. P. v. d. Heuvel, and O. Sporns. Rich Club Organization of Macaque Cerebral Cortex and Its Role in Network Communication. *PLoS One*, 7(9):e46497, 2012. doi: 10.1371/journal.pone.0046497.

D. E. Harrison and N. K. Larkin. El Niño-Southern Oscillation sea surface temperature and wind anomalies, 1946–1993. *Reviews of Geophysics*, 36(3):353–399, 1998. doi: 10.1029/98RG00715.

C. C. Hass. Social status in female bighorn sheep (Ovis canadensis): expression, development and reproductive correlates. *Journal of Zoology*, 225(3):509–523, 1991. doi: 10.1111/j.1469-7998.1991.tb03832.x.

J. Heitzig, J. F. Donges, Y. Zou, N. Marwan, and J. Kurths. Node-weighted measures for complex networks with spatially embedded, sampled, or differently sized nodes. *The European Physical Journal B - Condensed Matter and Complex Systems*, 85 (1):1–22, 2012. doi: 10.1140/epjb/e2011-20678-7.

I. M. Held, S. W. Lyons, and S. Nigam. Transients and the Extratropical Response to El Niño. *Journal of the Atmospheric Sciences*, 46(1):163–174, 1989. doi: 10.1175/1520-0469(1989)046<0163:TATERT>2.0.CO;2.

M. Helmstaedter, K. L. Briggman, S. C. Turaga, V. Jain, H. S. Seung, and W. Denk. Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, 500(7461):168–174, 2013. doi: 10.1038/nature12346.

H. H. Hendon, E. Lim, G. Wang, O. Alves, and D. Hudson. Prospects for predicting two flavors of El Niño. *Geophysical Research Letters*, 36(19):L19713, 2009. doi: 10.1029/2009GL040100.

C. Herrmann, M. Barthélemy, and P. Provero. Connectivity distribution of spatial networks. *Physical Review E*, 68(2):026128, 2003. doi: 10.1103/PhysRevE.68.026128.

J. Hlinka, D. Hartman, N. Jajcay, M. Vejmelka, R. Donner, N. Marwan, J. Kurths, and M. Paluš. Regional and inter-regional effects in evolving climate networks. *Nonlinear Processes in Geophysics*, 21(2):451–462, 2014. doi: 10.5194/npg-21-451-2014.

J. A. Hobson. *The Evolution of Modern Captitalism: A Study of Machine Production.* Scribners, 1906.

A. Hoell, C. Funk, T. Magadzire, J. Zinke, and G. Husak. El Niño–Southern Oscillation diversity and Southern Africa teleconnections during Austral Summer. *Climate Dynamics*, 45(5-6):1583–1599, 2015. doi: 10.1007/s00382-014-2414-z.

Z.-Z. Hu, A. Kumar, B. Jha, W. Wang, B. Huang, and B. Huang. An analysis of warm pool and cold tongue El Niños: air–sea coupling processes, global influences, and recent trends. *Climate Dynamics*, 38(9-10):2017–2035, 2011. doi: 10.1007/s00382-011-1224-9.

R. Jacob, K. P. Harikrishnan, R. Misra, and G. Ambika. Can recurrence networks show small-world property? *Physics Letters A*, 380(35):2718–2723, 2016. doi: 10.1016/j.physleta.2016.06.038.

T. A. Jarrell, Y. Wang, A. E. Bloniarz, C. A. Brittin, M. Xu, J. N. Thomson, D. G. Albertson, D. H. Hall, and S. W. Emmons. The Connectome of a Decision-Making Neural Network. *Science*, 337(6093):437–444, 2012. doi: 10.1126/science.1221762.

A. J. Jarvis, S. J. Jarvis, and C. N. Hewitt. Resource acquisition, distribution and end-use efficiencies and the growth of industrial society. *Earth System Dynamics*, 6(2):689–702, 2015. doi: 10.5194/esd-6-689-2015.

N. C. Johnson. How Many ENSO Flavors Can We Distinguish? *Journal of Climate*, 26(13):4816–4827, 2013. doi: 10.1175/JCLI-D-12-00649.1.

P. Jones, K. Trenberth, P. Ambenje, R. Bojariu, D. Easterling, T. Klein, D. Parker, J. Renwick, M. Rusticucci, B. Soden, and others. Observations: surface and atmospheric climate change. *Climate change 2007: the physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 235–336, 2007.

M. Kaiser and C. C. Hilgetag. Spatial growth of real-world networks. *Physical Review E*, 69(3), 2004. doi: 10.1103/PhysRevE.69.036103.

G. Kalna and D. J. Higham. Clustering Coefficients for Weighted Networks. In *Symposium on Network Analysis in Natural Sciences and Engineering*, page 45, 2006.

E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, R. Jenne, and D. Joseph. The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society*, 77(3):437–471, 1996. doi: 10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

H.-Y. Kao and J.-Y. Yu. Contrasting Eastern-Pacific and Central-Pacific Types of ENSO. *Journal of Climate*, 22(3):615–632, 2009. doi: 10.1175/2008JCLI2309.1.

T. R. Karl and K. E. Trenberth. Modern global climate change. *Science*, 302(5651): 1719–1723, 2003.

T. Karpouzoglou and J. Barron. A global and regional perspective of rainwater harvesting in sub-Saharan Africa's rainfed farming systems. *Physics and Chemistry of the Earth, Parts A/B/C*, 72–75:43–53, 2014. doi: 10.1016/j.pce.2014.09.009.

H.-M. Kim, P. J. Webster, and J. A. Curry. Impact of Shifting Patterns of Pacific Ocean Warming on North Atlantic Tropical Cyclones. *Science*, 325(5936):77–80, 2009. doi: 10.1126/science.1174062.

W. Kim, S.-W. Yeh, J.-H. Kim, J.-S. Kug, and M. Kwon. The unique 2009–2010 El Niño event: A fast phase transition of warm pool El Niño to La Niña. *Geophysical Research Letters*, 38(15):L15809, 2011. doi: 10.1029/2011GL048521.

R. Kinney, P. Crucitti, R. Albert, and V. Latora. Modeling cascading failures in the North American power grid. *The European Physical Journal B - Condensed Matter and Complex Systems*, 46(1):101–107, 2005. doi: 10.1140/epjb/e2005-00237-9.

R. Kistler, W. Collins, S. Saha, G. White, J. Woollen, E. Kalnay, M. Chelliah, W. Ebisuzaki, M. Kanamitsu, V. Kousky, H. van den Dool, R. Jenne, and M. Fiorino. The NCEP–NCAR 50–Year Reanalysis: Monthly Means CD–ROM and Documentation. *Bulletin of the American Meteorological Society*, 82(2):247–267, 2001. doi: 10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2.

M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter. Multilayer networks. *Journal of Complex Networks*, 2(3):203–271, 2014. doi: 10.1093/comnet/cnu016.

P. P. Klamser, M. Wiedermann, J. F. Donges, and R. V. Donner. Zealotry Effects on Opinion Dynamics in the Adaptive Voter Model. *preprint: arXiv:1612.06644*, 2016.

F. Klimm, J. Borge-Holthoefer, N. Wessel, J. Kurths, and G. Zamora-López. Individual node's contribution to the mesoscale of complex networks. *New Journal of Physics*, 16(12):125006, 2014. doi: 10.1088/1367-2630/16/12/125006.

E. V. Konstantinova, V. A. Skorobogatov, and M. V. Vidyuk. Applications of information theory in chemical graph theory. *Indian Journal of Chemistry*, 42(A): 1227–1240, 2003.

C. Korte and S. Milgram. Acquaintance networks between racial groups: Application of the small world method. *Journal of Personality and Social Psychology*, 15(2): 101–108, 1970.

M. Kretschmer, D. Coumou, J. F. Donges, and J. Runge. Using Causal Effect Networks to Analyze Different Arctic Drivers of Midlatitude Winter Circulation. *Journal of Climate*, 29(11):4069–4081, 2016. doi: 10.1175/JCLI-D-15-0654.1.

J.-S. Kug and Y.-G. Ham. Are there two types of La Niña? *Geophysical Research Letters*, 38(16):L16704, 2011. doi: 10.1029/2011GL048237.

J.-S. Kug, F.-F. Jin, and S.-I. An. Two Types of El Niño Events: Cold Tongue El Niño and Warm Pool El Niño. *Journal of Climate*, 22(6):1499–1515, 2009. doi: 10.1175/2008JCLI2624.1.

K. K. Kumar, B. Rajagopalan, M. Hoerling, G. Bates, and M. Cane. Unraveling the Mystery of Indian Monsoon Failure During El Niño. *Science*, 314(5796):115–119, 2006. doi: 10.1126/science.1131152.

Y. Kushnir, W. A. Robinson, I. Bladé, N. M. J. Hall, S. Peng, and R. Sutton. Atmospheric GCM Response to Extratropical SST Anomalies: Synthesis and Evaluation. *Journal of Climate*, 15(16):2233–2256, 2002. doi: 10.1175/1520-0442(2002)015<2233:AGRTES>2.0.CO;2.

P. W. Lamberti, M. T. Martin, A. Plastino, and O. A. Rosso. Intensive entropic non-triviality measure. *Physica A: Statistical Mechanics and its Applications*, 334 (1–2):119–131, 2004. doi: 10.1016/j.physa.2003.11.005.

A. Lancichinetti, M. Kivelä, J. Saramäki, and S. Fortunato. Characterizing the Community Structure of Complex Networks. *PLoS One*, 5(8):e11976, 2010. doi: 10.1371/journal.pone.0011976.

H. Lange, O. A. Rosso, and M. Hauhs. Ordinal pattern and statistical complexity analysis of daily stream flow time series. *The European Physical Journal Special Topics*, 222(2):535–552, 2013. doi: 10.1140/epjst/e2013-01858-3.

H. Lange and S. Boese. Recurrence Quantification and Recurrence Network Analysis of Global Photosynthetic Activity. In C. L. W. Jr and N. Marwan, editors, *Recurrence Quantification Analysis*, Understanding Complex Systems, pages 349–374. Springer International Publishing, 2015. ISBN 978-3-319-07154-1 978-3-319-07155-8. doi: 10.1007/978-3-319-07155-8_12.

N. K. Larkin and D. E. Harrison. Global seasonal temperature and precipitation anomalies during El Niño autumn and winter. *Geophysical Research Letters*, 32 (16):L16705, 2005. doi: 10.1029/2005GL022860.

V. Latora and M. Marchiori. Economic small-world behavior in weighted networks. *The European Physical Journal B - Condensed Matter and Complex Systems*, 32 (2):249–263, 2003. doi: 10.1140/epjb/e2003-00095-5.

M. Leblanc, S. Tweed, A. Van Dijk, and B. Timbal. A review of historic and future hydrological changes in the Murray-Darling Basin. *Global and Planetary Change*, 80–81:226–246, 2012. doi: 10.1016/j.gloplacha.2011.10.012.

M. Lenzen, K. Kanemoto, D. Moran, and A. Geschke. Mapping the Structure of the World Economy. *Environmental Science & Technology*, 46(15):8374–8381, 2012. doi: 10.1021/es300171x.

Q. Liu, N. Wen, and Z. Liu. An observational study of the impact of the North Pacific SST on the atmosphere. *Geophysical Research Letters*, 33(18):L18611, 2006. doi: 10.1029/2006GL026082.

R. López-Ruiz, H. L. Mancini, and X. Calbet. A statistical measure of complexity. *Physics Letters A*, 209(5):321–326, 1995. doi: 10.1016/0375-9601(95)00867-5.

C. Lorenz and H. Kunstmann. The Hydrological Cycle in Three State-of-the-Art Reanalyses: Intercomparison and Performance Analysis. *Journal of Hydrometeorology*, 13(5):1397–1420, 2012. doi: 10.1175/JHM-D-11-088.1.

D. F. Lott. Dominance Relations and Breeding Rate in Mature Male American Bison. *Zeitschrift für Tierpsychologie*, 49(4):418–432, 1979. doi: 10.1111/j.1439-0310.1979.tb00302.x.

J. Ludescher, A. Gozolchiani, M. I. Bogachev, A. Bunde, S. Havlin, and H. J. Schellnhuber. Improved El Niño forecasting by cooperativity detection. *Proceedings of the National Academy of Sciences*, 110(29):11742–11745, 2013. doi: 10.1073/pnas.1309353110.

J. Ludescher, A. Gozolchiani, M. I. Bogachev, A. Bunde, S. Havlin, and H. J. Schellnhuber. Very early warning of next El Niño. *Proceedings of the National Academy of Sciences*, 111(6):2064–2066, 2014. doi: 10.1073/pnas.1323058111.

D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson. The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*, 54(4):396–405, 2003. doi: 10.1007/s00265-003-0651-y.

A. C. Lute and J. T. Abatzoglou. Role of extreme snowfall events in interannual variability of snowfall accumulation in the western United States. *Water Resources Research*, 50(4):2874–2888, 2014. doi: 10.1002/2013WR014465.

N. Malik, N. Marwan, and J. Kurths. Spatial structures and directionalities in Monsoonal precipitation over South Asia. *Nonlinear Processes in Geophysics*, 17 (5):371–381, 2010. doi: 10.5194/npg-17-371-2010.

N. Malik, B. Bookhagen, N. Marwan, and J. Kurths. Analysis of spatial and temporal extreme monsoonal rainfall over South Asia using complex networks. *Climate Dynamics*, 39(3-4):971–987, 2011. doi: 10.1007/s00382-011-1156-4.

J. Maluck and R. V. Donner. A Network of Networks Perspective on Global Trade. *PLoS One*, 10(7):e0133310, 2015. doi: 10.1371/journal.pone.0133310.

N. T. Markov, M. M. Ercsey-Ravasz, R. Gomes, A. R, C. Lamy, L. Magrou, J. Vezoli, P. Misery, A. Falchier, R. Quilodran, M. A. Gariel, J. Sallet, R. Gamanut, C. Huissoud, S. Clavagnier, P. Giroud, D. Sappey-Marinier, P. Barone, C. Dehay, Z. Toroczkai, K. Knoblauch, V. Essen, D. C, and H. Kennedy. A Weighted and Directed Interareal Connectivity Matrix for Macaque Cerebral Cortex. *Cerebral Cortex*, 24(1):17–36, 2014a. doi: 10.1093/cercor/bhs270.

N. T. Markov, M. Ercsey-Ravasz, C. Lamy, A. R. R. Gomes, L. Magrou, P. Misery, P. Giroud, P. Barone, C. Dehay, Z. Toroczkai, K. Knoblauch, D. C. V. Essen, and H. Kennedy. The role of long-range connections on the specificity of the macaque interareal cortical network. *Proceedings of the National Academy of Sciences*, 110 (13):5187–5192, 2013. doi: 10.1073/pnas.1218972110.

N. T. Markov, J. Vezoli, P. Chameau, A. Falchier, R. Quilodran, C. Huissoud, C. Lamy, P. Misery, P. Giroud, S. Ullman, P. Barone, C. Dehay, K. Knoblauch, and H. Kennedy. Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex. *Journal of Comparative Neurology*, 522(1):225–259, 2014b. doi: 10.1002/cne.23458.

E. A. Martin, M. Paczuski, and J. Davidsen. Interpretation of link fluctuations in climate networks during El Niño periods. *EPL (Europhysics Letters)*, 102(4): 48003, 2013.

M. T. Martin, A. Plastino, and O. A. Rosso. Statistical complexity and disequilibrium. *Physics Letters A*, 311(2–3):126–132, 2003. doi: 10.1016/S0375-9601(03)00491-2.

M. Martin, A. Plastino, and O. Rosso. Generalized statistical complexity measures: Geometrical and analytical properties. *Physica A: Statistical Mechanics and its Applications*, 369(2):439–462, 2006.

N. Marwan, J. F. Donges, Y. Zou, R. V. Donner, and J. Kurths. Complex network approach for recurrence analysis of time series. *Physics Letters A*, 373(46):4246–4254, 2009. doi: 10.1016/j.physleta.2009.09.042.

S. Maslov and K. Sneppen. Specificity and Stability in Topology of Protein Networks. *Science*, 296(5569):910–913, 2002. doi: 10.1126/science.1065103.

S. Maslov, K. Sneppen, and A. Zaliznyak. Detection of topological patterns in complex networks: correlation profile of the internet. *Physica A: Statistical Mechanics and its Applications*, 333:529–540, 2004. doi: 10.1016/j.physa.2003.06.002.

M. P. McCormick, L. W. Thomason, and C. R. Trepte. Atmospheric effects of the Mt Pinatubo eruption. *Nature*, 373(6513):399–404, 1995. doi: 10.1038/373399a0.

P. K. McGregor. *Animal communication networks*. Cambridge University Press, 2005.

P. J. Menck, J. Heitzig, J. Kurths, and H. Joachim Schellnhuber. How dead ends undermine power grid stability. *Nature Communications*, 5, 2014. doi: 10.1038/ncomms4969.

S. Milgram. The small world problem. *Psychology Today*, 2(1):60–67, 1967.

R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon. Superfamilies of evolved and designed networks. *Science*, 303(5663): 1538–1542, 2004.

N. Molkenthin, K. Rehfeld, N. Marwan, and J. Kurths. Networks from Flows - From Dynamics to Topology. *Scientific Reports*, 4, 2014. doi: 10.1038/srep04119.

M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms*, 6(2-3):161–180, 1995. doi: 10.1002/rsa.3240060204.

J. D. Neelin, C. Chou, and H. Su. Tropical drought regions in global warming and El Niño teleconnections. *Geophysical Research Letters*, 30(24):2275, 2003. doi: 10.1029/2003GL018625.

M. Newman. The Structure and Function of Complex Networks. *SIAM Review*, 45 (2):167–256, 2003. doi: 10.1137/S003614450342480.

M. E. J. Newman. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*, 64(1):016132, 2001. doi: 10.1103/PhysRevE.64.016132.

M. E. J. Newman. A measure of betweenness centrality based on random walks. *Social Networks*, 27(1):39–54, 2005. doi: 10.1016/j.socnet.2004.11.009.

M. E. J. Newman, S. Forrest, and J. Balthrop. Email networks and the spread of computer viruses. *Physical Review E*, 66(3):035101, 2002.

M. Newman. *Networks: An Introduction*. OUP Oxford, 2010. ISBN 978-0-19-150070-1.

S. Nicholson and J. Selato. The influence of La Niña on African rainfall. *International Journal of Climatology*, 20(14):1761–1776, 2000. doi: 10.1002/1097-0088(20001130)20:14<1761::AID-JOC580>3.0.CO;2-W.

G. R. North, T. L. Bell, R. F. Cahalan, and F. J. Moeng. Sampling Errors in the Estimation of Empirical Orthogonal Functions. *Monthly Weather Review*, 110(7): 699–706, 1982. doi: 10.1175/1520-0493(1982)110<0699:SEITEO>2.0.CO;2.

S. W. Oh, J. A. Harris, L. Ng, B. Winslow, N. Cain, S. Mihalas, Q. Wang, C. Lau, L. Kuan, A. M. Henry, M. T. Mortrud, B. Ouellette, T. N. Nguyen, S. A. Sorensen, C. R. Slaughterbeck, W. Wakeman, Y. Li, D. Feng, A. Ho, E. Nicholas, K. E. Hirokawa, P. Bohn, K. M. Joines, H. Peng, M. J. Hawrylycz, J. W. Phillips, J. G. Hohmann, P. Wohnoutka, C. R. Gerfen, C. Koch, A. Bernard, C. Dang, A. R. Jones, and H. Zeng. A mesoscale connectome of the mouse brain. *Nature*, 508 (7495):207–214, 2014. doi: 10.1038/nature13186.

T. Opsahl. Why Anchorage is not (that) important: Binary ties and Sample selection, 2011. Published: www.toreopsahl.com/2011/08/12/.

M. Paluš, D. Hartman, J. Hlinka, and M. Vejmelka. Discerning connectivity from dynamics in climate networks. *Nonlinear Processes in Geophysics*, 18(5):751–763, 2011. doi: 10.5194/npg-18-751-2011.

G. Parisi. Statistical Physics and biology. *Physics World*, 6(9):42, 1993. doi: 10.1088/2058-7058/6/9/35.

M. Penrose. *Random geometric graphs. Vol. 5.* Oxford University Press, Oxford, 2003. ISBN 0-19-850626-0.

O. L. Phillips, L. E. O. C. Aragão, S. L. Lewis, J. B. Fisher, J. Lloyd, G. López-González, Y. Malhi, A. Monteagudo, J. Peacock, C. A. Quesada, G. van der Heijden, S. Almeida, I. Amaral, L. Arroyo, G. Aymard, T. R. Baker, O. Bánki, L. Blanc, D. Bonal, P. Brando, J. Chave, Á. C. A. de Oliveira, N. D. Cardozo, C. I. Czimczik, T. R. Feldpausch, M. A. Freitas, E. Gloor, N. Higuchi, E. Jiménez, G. Lloyd, P. Meir, C. Mendoza, A. Morel, D. A. Neill, D. Nepstad, S. Patiño, M. C. Peñuela, A. Prieto, F. Ramírez, M. Schwarz, J. Silva, M. Silveira, A. S. Thomas, H. t. Steege, J. Stropp, R. Vásquez, P. Zelazowski, E. A. Dávila, S. Andelman, A. Andrade, K.-J. Chao, T. Erwin, A. Di Fiore, E. H. C., H. Keeling, T. J. Killeen, W. F. Laurance, A. P. Cruz, N. C. A. Pitman, P. N. Vargas, H. Ramírez-Angulo, A. Rudas, R. Salamão, N. Silva, J. Terborgh, and A. Torres-Lezama. Drought Sensitivity of the Amazon Rainforest. *Science*, 323(5919):1344–1347, 2009. doi: 10.1126/science.1164033.

D. Pozo-Vázquez, S. R. Gámiz-Fortis, J. Tovar-Pescador, M. J. Esteban-Parra, and Y. Castro-Díez. El Niño–southern oscillation events and associated European winter precipitation anomalies. *International Journal of Climatology*, 25(1):17–31, 2005. doi: 10.1002/joc.1097.

W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C*, volume 2. 1996.

A. Radebach, R. V. Donner, J. Runge, J. F. Donges, and J. Kurths. Disentangling different types of El Niño episodes by evolving climate network analysis. *Physical Review E*, 88(5):052807, 2013. doi: 10.1103/PhysRevE.88.052807.

A. Rammig, M. Wiedermann, J. F. Donges, F. Babst, W. von Bloh, D. Frank, K. Thonicke, and M. D. Mahecha. Coincidences of climate extremes and anomalous vegetation responses: comparing tree ring patterns to simulated productivity. *Biogeosciences*, 12(2):373–385, 2015. doi: 10.5194/bg-12-373-2015.

N. Rashevsky. Life, information theory, and topology. *The Bulletin of mathematical biophysics*, 17(3):229–235, 1955. doi: 10.1007/BF02477860.

E. M. Rasmusson and T. H. Carpenter. Variations in Tropical Sea Surface Temperature and Surface Wind Fields Associated with the Southern Oscillation/El Niño. *Monthly Weather Review*, 110(5):354–384, 1982. doi: 10.1175/1520-0493(1982)110<0354:VITSST>2.0.CO;2.

E. Rataj, K. Kunzweiler, and S. Garthus-Niegel. Extreme weather events in developing countries and related injuries and mental health disorders-a systematic review. *BMC public health*, 16(1):1020, 2016.

E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási. Hierarchical Organization of Modularity in Metabolic Networks. *Science*, 297(5586): 1551–1555, 2002. doi: 10.1126/science.1073374.

E. Ravasz and A.-L. Barabási. Hierarchical organization in complex networks. *Physical Review E*, 67(2):026112, 2003. doi: 10.1103/PhysRevE.67.026112.

N. A. Rayner, D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres*, 108(D14):4407, 2003. doi: 10.1029/2002JD002670.

H.-L. Ren and F.-F. Jin. Niño indices for two types of ENSO. *Geophysical Research Letters*, 38(4):L04704, 2011. doi: 10.1029/2010GL046031.

M. A. d. Reus and M. P. v. d. Heuvel. Rich Club Organization and Intermodule Communication in the Cat Connectome. *Journal of Neuroscience*, 33(32):12929–12939, 2013. doi: 10.1523/JNEUROSCI.1448-13.2013.

A. Rheinwalt, N. Marwan, J. Kurths, P. Werner, and F.-W. Gerstengarbe. Boundary effects in network measures of spatially embedded networks. *Europhysics Letters*, 100(2):28002, 2012. doi: 10.1209/0295-5075/100/28002.

J. C. Rogers. Precipitation Variability over the Caribbean and Tropical Americas Associated with the Southern Oscillation. *Journal of Climate*, 1(2):172–182, 1988. doi: 10.1175/1520-0442(1988)001<0172:PVOTCA>2.0.CO;2.

C. F. Ropelewski and M. S. Halpert. Global and Regional Scale Precipitation Patterns Associated with the El Niño/Southern Oscillation. *Monthly Weather Review*, 115(8):1606–1626, 1987. doi: 10.1175/1520-0493(1987)115<1606:GARSPP>2.0.CO;2.

C. F. Ropelewski and M. S. Halpert. Quantifying Southern Oscillation-Precipitation Relationships. *Journal of Climate*, 9(5):1043–1059, 1996. doi: 10.1175/1520-0442(1996)009<1043:QSOPR>2.0.CO;2.

O. E. Rössler. An equation for continuous chaos. *Physics Letters A*, 57(5):397–398, 1976. doi: 10.1016/0375-9601(76)90101-8.

O. A. Rosso, H. A. Larrondo, M. T. Martin, A. Plastino, and M. A. Fuentes. Distinguishing Noise from Chaos. *Physical Review Letters*, 99(15), 2007. doi: 10.1103/PhysRevLett.99.154102.

M. Rubinov, R. J. F. Ypma, C. Watson, and E. T. Bullmore. Wiring cost and topological participation of the mouse brain connectome. *Proceedings of the National Academy of Sciences*, 112(32):10032–10037, 2015. doi: 10.1073/pnas.1420315112.

B. Rudolf, H. Hauschild, W. Rueth, and U. Schneider. Terrestrial Precipitation Analysis: Operational Method and Required Density of Point Measurements. In M. Desbois and F. Désalmand, editors, *Global Precipitations and Climate Change*, number 26 in NATO ASI Series, pages 173–186. Springer Berlin Heidelberg, 1994. ISBN 978-3-642-79270-0 978-3-642-79268-7. doi: 10.1007/978-3-642-79268-7_10.

J. Runge, J. Heitzig, N. Marwan, and J. Kurths. Quantifying causal coupling strength: A lag-specific measure for multivariate time series related to transfer entropy. *Physical Review E*, 86(6):061121, 2012a. doi: 10.1103/PhysRevE.86.061121.

J. Runge, J. Heitzig, V. Petoukhov, and J. Kurths. Escaping the Curse of Dimensionality in Estimating Multivariate Transfer Entropy. *Physical Review Letters*, 108(25):258701, 2012b. doi: 10.1103/PhysRevLett.108.258701.

J. Runge, V. Petoukhov, and J. Kurths. Quantifying the Strength and Delay of Climatic Interactions: The Ambiguities of Cross Correlation and a Novel Measure Based on Graphical Models. *Journal of Climate*, 27(2):720–739, 2013. doi: 10.1175/JCLI-D-13-00159.1.

J. Runge, V. Petoukhov, J. F. Donges, J. Hlinka, N. Jajcay, M. Vejmelka, D. Hartman, N. Marwan, M. Paluš, and J. Kurths. Identifying causal gateways and mediators in complex spatio-temporal systems. *Nature Communications*, 6:8502, 2015. doi: 10.1038/ncomms9502.

D. S. Sade. Sociometrics of Macaca mulatta I. Linkages and Cliques in Grooming Matrices. *Folia Primatologica*, 18(3-4):196–223, 2004. doi: 10.1159/000155480.

J. Saramäki, M. Kivelä, J.-P. Onnela, K. Kaski, and J. Kertész. Generalizations of the clustering coefficient to weighted complex networks. *Physical Review E*, 75(2):027105, 2007. doi: 10.1103/PhysRevE.75.027105.

M. W. Schein and M. H. Fohrman. Social dominance relationships in a herd of dairy cattle. *The British Journal of Animal Behaviour*, 3(2):45–55, 1955. doi: 10.1016/S0950-5601(55)80012-3.

C.-F. Schleussner, J. F. Donges, R. V. Donner, and H. J. Schellnhuber. Armed-conflict risks enhanced by climate-related disasters in ethnically fractionalized

countries. *Proceedings of the National Academy of Sciences*, 113(33):9216–9221, 2016. doi: 10.1073/pnas.1601611113.

U. Schneider, A. Becker, P. Finger, A. Meyer-Christoffer, B. Rudolf, and M. Ziese. *GPCC Full Data Reanalysis Version 7.0 (at 0.5°, 1.0°, 2.5°): Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data*. Global Precipitation Climatology Centre (GPCC, http://gpcc.dwd.de/) at Deutscher Wetterdienst, 2015.

T. Schreiber and A. Schmitz. Surrogate time series. *Physica D: Nonlinear Phenomena*, 142(3–4):346–382, 2000. doi: 10.1016/S0167-2789(00)00043-9.

P. Schultz, J. Heitzig, and J. Kurths. A random growth model for power grids and other spatially embedded infrastructure networks. *The European Physical Journal Special Topics*, 223(12):2593–2610, 2014. doi: 10.1140/epjst/e2014-02279-6.

D. W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. John Wiley & Sons, 2015. ISBN 978-1-118-57553-6.

M. C. Serreze and J. A. Francis. The Arctic Amplification Debate. *Climatic Change*, 76(3-4):241–264, 2006. doi: 10.1007/s10584-005-9017-y.

J. Shaman and E. Tziperman. An Atmospheric Teleconnection Linking ENSO and Southwestern European Precipitation. *Journal of Climate*, 24(1):124–139, 2010. doi: 10.1175/2010JCLI3590.1.

J. F. Siegmund, M. Wiedermann, J. F. Donges, and R. V. Donner. Impact of temperature and precipitation extremes on the flowering dates of four German wildlife shrub species. *Biogeosciences*, 13(19):5541–5555, 2016a. doi: 10.5194/bg-13-5541-2016.

J. F. Siegmund, T. G. M. Sanders, I. Heinrich, E. van der Maaten, S. Simard, G. Helle, and R. V. Donner. Meteorological Drivers of Extremes in Daily Stem Radius Variations of Beech, Oak, and Pine in Northeastern Germany: An Event Coincidence Analysis. *Frontiers in Plant Science*, 7, 2016b. doi: 10.3389/fpls.2016.00733.

J. F. Siegmund, N. Siegmund, and R. V. Donner. CoinCalc - A new R package for quantifying simultaneities of event series. *Computers & Geosciences*, 98:64–72, 2017. doi: 10.1016/j.cageo.2016.10.004.

M. Small. Complex networks from time series: Capturing dynamics. pages 2509–2512. IEEE, 2013. ISBN 978-1-4673-5762-3 978-1-4673-5760-9 978-1-4673-5761-6. doi: 10.1109/ISCAS.2013.6572389.

D. J. A. Smit, C. J. Stam, D. Posthuma, D. I. Boomsma, and E. J. C. de Geus. Heritability of "small-world" networks in the brain: A graph theoretical analysis of resting-state EEG functional connectivity. *Human Brain Mapping*, 29(12):1368–1378, 2008. doi: 10.1002/hbm.20468.

H. F. Song, H. Kennedy, and X.-J. Wang. Spatial embedding of structural similarity in the cerebral cortex. *Proceedings of the National Academy of Sciences*, 111(46): 16580–16585, 2014. doi: 10.1073/pnas.1414153111.

O. Sporns, G. Tononi, and R. Kötter. The human connectome: a structural description of the human brain. *PLoS Computational Biology*, 1(4):e42, 2005.

K. Steinhaeuser, A. R. Ganguly, and N. V. Chawla. Multivariate and multiscale dependence in the global climate system revealed through complex networks. *Climate Dynamics*, 39(3-4):889–895, 2011. doi: 10.1007/s00382-011-1135-9.

V. Stolbova, P. Martin, B. Bookhagen, N. Marwan, and J. Kurths. Topology and seasonal evolution of the network of extreme precipitation over the Indian subcontinent and Sri Lanka. *Nonlinear Processes in Geophysics*, 21(4):901–917, 2014. doi: 10.5194/npg-21-901-2014.

V. Stolbova, E. Surovyatkina, B. Bookhagen, and J. Kurths. Tipping elements of the Indian monsoon: Prediction of onset and withdrawal. *Geophysical Research Letters*, 43(8):3982–3990, 2016. doi: 10.1002/2016GL068392.

H. v. Storch and F. W. Zwiers. *Statistical Analysis in Climate Research*. Cambridge University Press, 2001. ISBN 978-0-521-01230-0.

S. R. Sundaresan, I. R. Fischhoff, J. Dushoff, and D. I. Rubenstein. Network metrics reveal differences in social organization between two fission–fusion species, Grevy's zebra and onager. *Oecologia*, 151(1):140–149, 2007. doi: 10.1007/s00442-006-0553-6.

A. Tantet and H. A. Dijkstra. An interaction network perspective on the relation between patterns of sea surface temperature variability and global mean surface temperature. *Earth System Dynamics*, 5(1):1–14, 2014. doi: 10.5194/esd-5-1-2014.

A. S. Taschetto and M. H. England. El Niño Modoki Impacts on Australian Rainfall. *Journal of Climate*, 22(11):3167–3174, 2009. doi: 10.1175/2008JCLI2589.1.

R. G. Tedeschi, I. F. A. Cavalcanti, and A. M. Grimm. Influences of two types of ENSO on South American precipitation. *International Journal of Climatology*, 33 (6):1382–1400, 2013. doi: 10.1002/joc.3519.

J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. Doyne Farmer. Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena*, 58(1):77–94, 1992. doi: 10.1016/0167-2789(92)90102-S.

G. Tirabassi and C. Masoller. On the effects of lag-times in networks constructed from similarities of monthly fluctuations of climate fields. *Europhysics Letters*, 102 (5):59003, 2013. doi: 10.1209/0295-5075/102/59003.

G. Tirabassi, C. Masoller, and M. Barreiro. A study of the air-sea interaction in the South Atlantic Convergence Zone through Granger causality: AIR-SEA INTERACTION IN THE SACZ THROUGH GRANGER CAUSALITY. *International Journal of Climatology*, 35(12):3440–3453, 2015. doi: 10.1002/joc.4218.

K. E. Trenberth. Signal Versus Noise in the Southern Oscillation. *Monthly Weather Review*, 112(2):326–332, 1984. doi: 10.1175/1520-0493(1984)112<0326:SVNITS>2.0.CO;2.

K. E. Trenberth. The Definition of El Niño. *Bulletin of the American Meteorological Society*, 78(12):2771–2777, 1997. doi: 10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2.

K. E. Trenberth and J. W. Hurrell. Decadal atmosphere-ocean variations in the Pacific. *Climate Dynamics*, 9(6):303–319, 1994. doi: 10.1007/BF00204745.

A. A. Tsonis and P. J. Roebber. The architecture of the climate network. *Physica A: Statistical Mechanics and its Applications*, 333:497–504, 2004. doi: 10.1016/j.physa.2003.10.045.

A. A. Tsonis and K. L. Swanson. Topology and Predictability of El Niño and La Niña Networks. *Physical Review Letters*, 100(22):228502, 2008. doi: 10.1103/PhysRevLett.100.228502.

A. A. Tsonis, K. L. Swanson, and P. J. Roebber. What Do Networks Have to Do with Climate? *Bulletin of the American Meteorological Society*, 87(5):585–595, 2006. doi: 10.1175/BAMS-87-5-585.

A. A. Tsonis, K. L. Swanson, and G. Wang. On the Role of Atmospheric Teleconnections in Climate. *Journal of Climate*, 21(12):2990–3001, 2008. doi: 10.1175/2007JCLI1907.1.

A. A. Tsonis, G. Wang, K. L. Swanson, F. A. Rodrigues, and L. d. F. Costa. Community structure and dynamics in climate networks. *Climate Dynamics*, 37(5-6):933–940, 2010. doi: 10.1007/s00382-010-0874-3.

L. Tupikina, K. Rehfeld, N. Molkenthin, V. Stolbova, N. Marwan, and J. Kurths. Characterizing the evolution of climate networks. *Nonlinear Processes in Geophysics*, 21(3):705–711, 2014. doi: 10.5194/npg-21-705-2014.

S. M. Uppala, P. W. Kållberg, A. J. Simmons, U. Andrae, V. D. C. Bechtold, M. Fiorino, J. K. Gibson, J. Haseler, A. Hernandez, G. A. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R. P. Allan, E. Andersson, K. Arpe, M. A. Balmaseda, A. C. M. Beljaars, L. V. D. Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Hólm, B. J. Hoskins, L. Isaksen, P. a. E. M. Janssen, R. Jenne, A. P. Mcnally, J.-F. Mahfouf, J.-J. Morcrette, N. A. Rayner, R. W. Saunders, P. Simon, A. Sterl, K. E. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo, and J. Woollen. The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, 131(612):2961–3012, 2005. doi: 10.1256/qj.04.176.

M. van der Mheen, H. A. Dijkstra, A. Gozolchiani, M. den Toom, Q. Feng, J. Kurths, and E. Hernandez-Garcia. Interaction network based early warning indicators for the Atlantic MOC collapse. *Geophysical Research Letters*, 40(11):2714–2719, 2013. doi: 10.1002/grl.50515.

A. I. J. M. van Dijk, H. E. Beck, R. S. Crosbie, R. A. M. de Jeu, Y. Y. Liu, G. M. Podger, B. Timbal, and N. R. Viney. The Millennium Drought in southeast Australia (2001–2009): Natural and human causes and implications for water resources, ecosystems, economy, and society. *Water Resources Research*, 49(2): 1040–1057, 2013. doi: 10.1002/wrcr.20123.

L. R. Varshney, B. L. Chen, E. Paniagua, D. H. Hall, and D. B. Chklovskii. Structural Properties of the Caenorhabditis elegans Neuronal Network. *PLoS Computational Biology*, 7(2):e1001066, 2011. doi: 10.1371/journal.pcbi.1001066.

P. E. Vertes, A. F. Alexander-Bloch, N. Gogtay, J. N. Giedd, J. L. Rapoport, and E. T. Bullmore. Simple models of human brain functional networks. *Proceedings of the National Academy of Sciences*, 109(15):5868–5873, 2012. doi: 10.1073/pnas.1111738109.

A. Vespignani. Complex networks: The fragility of interdependency. *Nature*, 464 (7291):984–985, 2010. doi: 10.1038/464984a.

R. Wackerbauer, A. Witt, H. Atmanspacher, J. Kurths, and H. Scheingraber. A comparative classification of complexity measures. *Chaos, Solitons & Fractals*, 4 (1):133–173, 1994. doi: 10.1016/0960-0779(94)90023-X.

B. Wang, H. Tang, C. Guo, and Z. Xiu. Entropy optimization of scale-free networks' robustness to random failures. *Physica A: Statistical Mechanics and its Applications*, 363(2):591–596, 2006. doi: 10.1016/j.physa.2005.08.025.

J. Wang, L. Wang, Y. Zang, H. Yang, H. Tang, Q. Gong, Z. Chen, C. Zhu, and Y. He. Parcellation-dependent small-world brain functional networks: A resting-state fMRI study. *Human Brain Mapping*, 30(5):1511–1523, 2009. doi: 10.1002/hbm.20623.

Y. Wang, A. Gozolchiani, Y. Ashkenazy, and S. Havlin. Oceanic El-Niño wave dynamics and climate networks. *New Journal of Physics*, 18(3):033021, 2016. doi: 10.1088/1367-2630/18/3/033021.

D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998. doi: 10.1038/30918.

N. Wen, Z. Liu, Q. Liu, and C. Frankignoul. Observations of SST, heat flux and North Atlantic Ocean-atmosphere interaction. *Geophysical Research Letters*, 32 (24):L24619, 2005. doi: 10.1029/2005GL024871.

D. R. White and M. Houseman. The navigability of strong ties: Small worlds, tie strength, and network topology. *Complexity*, 8(1):72–81, 2002.

M. Wiedermann, J. F. Donges, J. Heitzig, and J. Kurths. Node-weighted interacting network measures improve the representation of real-world complex systems. *Europhysics Letters*, 102(2):28007, 2013. doi: 10.1209/0295-5075/102/28007.

M. Wiedermann, J. F. Donges, J. Heitzig, W. Lucht, and J. Kurths. Macroscopic description of complex adaptive networks coevolving with dynamic node states. *Physical Review E*, 91(5):052801, 2015. doi: 10.1103/PhysRevE.91.052801.

M. Wiedermann, J. F. Donges, D. Handorf, J. Kurths, and R. V. Donner. Hierarchical structures in Northern Hemispheric extratropical winter ocean-atmosphere interactions. *International Journal of Climatology*, 2016a. doi: 10.1002/joc.4956.

M. Wiedermann, J. F. Donges, J. Kurths, and R. V. Donner. Spatial network surrogates for disentangling complex system structure from spatial embedding of nodes. *Physical Review E*, 93(4):042308, 2016b. doi: 10.1103/PhysRevE.93.042308.

M. Wiedermann, A. Radebach, J. F. Donges, J. Kurths, and R. V. Donner. A climate network-based index to discriminate different types of El Niño and La Niña. *Geophysical Research Letters*, 43(13):2016GL069119, 2016c. doi: 10.1002/2016GL069119.

M. Wiedermann, J. F. Donges, J. Kurths, and R. V. Donner. Mapping and discrimination of networks in the complexity-entropy plane. *preprint: arXiv:1704.07599*, 2017a.

M. Wiedermann, J. F. Siegmund, J. F. Donges, J. Kurths, and R. V. Donner. Differential imprints of distinct ENSO flavors in global extreme precipitation patterns. *preprint: arXiv:1702.00218*, 2017b.

T. Woollings, A. Hannachi, and B. Hoskins. Variability of the North Atlantic eddy-driven jet stream. *Quarterly Journal of the Royal Meteorological Society*, 136(649): 856–868, 2010. doi: 10.1002/qj.625.

K. Wyrtki. El Niño – The Dynamic Response of the Equatorial Pacific Oceanto Atmospheric Forcing. *Journal of Physical Oceanography*, 5(4):572–584, 1975. doi: 10.1175/1520-0485(1975)005<0572:ENTDRO>2.0.CO;2.

K. Yamasaki, A. Gozolchiani, and S. Havlin. Climate Networks around the Globe are Significantly Affected by El Niño. *Physical Review Letters*, 100(22):228501, 2008. doi: 10.1103/PhysRevLett.100.228501.

S.-W. Yeh, J.-S. Kug, B. Dewitte, M.-H. Kwon, B. P. Kirtman, and F.-F. Jin. El Niño in a changing climate. *Nature*, 461(7263):511–514, 2009. doi: 10.1038/nature08316.

Y. Yuan and H. Yan. Different types of La Niña events and different responses of the tropical atmosphere. *Chinese Science Bulletin*, 58(3):406–415, 2012. doi: 10.1007/s11434-012-5423-5.

G. Zamora-López, V. Zlatić, C. Zhou, H. Štefančić, and J. Kurths. Reciprocity of networks with degree correlations and arbitrary degree sequences. *Physical Review E*, 77(1):016106, 2008. doi: 10.1103/PhysRevE.77.016106.

D. C. Zemp, M. Wiedermann, J. Kurths, A. Rammig, and J. F. Donges. Node-weighted measures for complex networks with directed and weighted edges for studying continental moisture recycling. *Europhysics Letters*, 107(5):58005, 2014. doi: 10.1209/0295-5075/107/58005.

B. Zhang and S. Horvath. A General Framework for Weighted Gene Co-Expression Network Analysis. *Statistical Applications in Genetics and Molecular Biology*, 4 (1), 2005. doi: 10.2202/1544-6115.1128.

R. Zhang, A. Sumi, and M. Kimoto. Impact of El Niño on the East Asian Monsoon. *Journal of the Meteorological Society of Japan. Ser. II*, 74(1):49–62, 1996. doi: 10.2151/jmsj1965.74.1_49.

R. Zhang, A. Sumi, and M. Kimoto. A diagnostic study of the impact of El Niño on the precipitation in China. *Advances in Atmospheric Sciences*, 16(2):229–241, 1999. doi: 10.1007/BF02973084.

C. Zhou, L. Zemanová, G. Zamora-López, C. C. Hilgetag, and J. Kurths. Structure–function relationship in complex brain networks expressed by hierarchical synchronization. *New Journal of Physics*, 9(6):178, 2007. doi: 10.1088/1367-2630/9/6/178.

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 30. Januar 2018 Marc Wiedermann